

# How Compressible are Sparse Innovation Processes?

Hamid Ghourchian, Arash Amini, *Senior Member, IEEE*, and  
Amin Gohari, *Senior Member, IEEE*

## Abstract

The sparsity and compressibility of finite-dimensional signals are of great interest in fields such as compressed sensing. The notion of compressibility is also extended to infinite sequences of i.i.d. or ergodic random variables based on the observed error in their nonlinear  $k$ -term approximation. In this work, we use the entropy measure to study the compressibility of continuous-domain innovation processes (alternatively known as white noise). Specifically, we define such a measure as the entropy limit of the doubly quantized (time and amplitude) process by identifying divergent terms and extracting the convergent part. While the converging part determines the introduced entropy, the diverging terms provide a tool to compare the compressibility of various innovation processes. In particular, we study stable, and impulsive Poisson innovation processes representing various type of distributions. Our results recognize Poisson innovations as the most compressible one with an entropy measure far below that of stable innovations. While this result departs from the previous knowledge regarding the compressibility of fat-tailed distributions, our entropy measure ranks stable innovations according to their tail decay.

## Index Terms

Compressibility, entropy, Impulsive Poisson process, stable innovation, white Lévy noise.

## I. INTRODUCTION

The compressible signal models have been extensively used to represent or approximate various types of data such as audio, image and video signals. Intuitively, a signal is called compressible if in its representation using a known dictionary only a few atoms contribute significantly and the rest amount to negligible contribution. Sparse signals are among the special

The authors are with the department of Electrical Engineering and Advanced Communication Research Institute (ACRI), Sharif University of Technology, Tehran, Iran. E-mails: h\_ghourchian@ee.sharif.edu, {aamini, aminzadeh}@sharif.ir.

cases for which the mentioned representation consists of a few non-zero (instead of insignificant) contributions. Compressible signals in general, and sparse signals in particular, are of fundamental importance in fields such as compressed sensing [1], [2], dimensionality reduction [3], and nonlinear approximation theory [4].

Traditionally, compressible signals are defined as infinite sequences within the Besov spaces [4], where the decay rate of the  $k$ -term approximation error could be efficiently controlled. A more recent deterministic approach towards modeling compressibility is via weak- $\ell_p$  spaces [2], [5]. The latter approach is useful in compressed sensing, where  $\ell_1$  (or  $\ell_p$ ) regularization techniques are used.

The study of stochastic models for compressibility started with identifying compressible priors. For this purpose, independent and identically distributed (i.i.d.) sequences of random variables with a given probability law are examined. Cevher in [6] defined the compressibility criterion based on the decay rate of the mean values of order statistics. A more precise definition in [7] revealed the connection between compressibility and the decay rate of the tail probability. In particular, heavy-tailed priors with infinite  $p$ -order moments were identified as  $\ell_p$ -compressible probability laws. It was later shown in [8] that this sufficient condition is indeed, necessary as well. A similar identification of heavy-tailed priors (with infinite variance) was obtained in [9] with a different definition of compressibility<sup>1</sup>.

The first non-i.i.d. result appeared in [10]. By extending the techniques used in [8], and based on the notion of  $\ell_p$ -compressibility of [7], it is shown in [10] that discrete-domain stationary and ergodic processes are  $\ell_p$ -compressible if and only if the invariant distribution of the process is an  $\ell_p$ -compressible prior.

The recent framework of sparse stochastic processes introduced in [11]–[13] extends the discrete-domain models to continuous-domain. In practice, most of the compressible discrete-domain signals arise from discretized versions of continuous-domain physical phenomena. Thus, it might be beneficial to have continuous-domain models that result in compressible/sparse discrete-domain models for a general class of sampling strategies. Indeed, this goal is achieved in [11], [12] by considering non-Gaussian stochastic processes. The building block of these models are the innovation processes (widely known as white noise) that mimic i.i.d. sequences

<sup>1</sup>Based on the definition in [9], a prior is called compressible if the compressed measurements (by applying a Gaussian ensemble) of a high-dimensional vector of i.i.d. values following this law could be better recovered using  $\ell_1$  regularization techniques than the classical  $\ell_2$  minimization approaches.

in continuous-domain. Unlike sequences, the probability laws of innovation processes are bound to a specific family known as *infinitely divisible*, that includes  $\alpha$ -stable ( $\alpha = 2$  corresponds to Gaussians) and compound Poisson distributions.

The discretization of innovation processes are known to form stationary and ergodic sequences of random variables with infinitely divisible distributions. As the tail probability of all non-Gaussian infinitely divisible laws are slower than Gaussians [14], [15], they exhibit more compressible behavior than Gaussians according to [7], [10].

In this paper, we examine the compressibility of innovation processes using the more traditional measure of entropy. In information theory, entropy naturally arises as the proper measure of compression for Shannon's lossless source coding problem. It also finds a geometrical interpretation as the volume of the typical sets. As the definition of entropy ignores the amplitude distribution of the involved random variables and only takes into account the distribution (or concentration) of the probability measure, it provides a fundamentally different perspective of compressibility compared to the previously studied  $k$ -term approximation. In particular, based on the entropy measure, we will show that impulsive Poisson innovation processes are by far more compressible than heavy-tailed  $\alpha$ -stable innovation processes; the previous studies on their  $k$ -term approximation sort them in the opposite order.

The two main challenges that are addressed in this paper are 1) defining an entropy measure for continuous-domain innovation processes that translates into operational lossless source coding, and 2) evaluating such a measure for particular instances to allow for their comparison. We recall that the *differential entropy* of a random variable  $X$  with continuous range is defined by finely quantizing  $X$  with resolution  $1/m$ , followed by canceling a diverging term  $\log(m)$  from the discrete entropy of its quantized version. Obviously, we shall expect more elaborate diverging terms when dealing with continuous-domain processes. More specifically, after appropriate quantization in time and amplitude with resolutions  $1/n$  and  $1/m$  respectively, we propose the following expression to cancel out the diverging terms:

$$\frac{\mathcal{H}_{m,n}(X)}{\kappa(n)} - \log(m) - \zeta(n), \quad (1)$$

where  $\mathcal{H}_{m,n}(X)$  is the discrete entropy of the time/amplitude quantized process, and  $\kappa(\cdot)$  and  $\zeta(\cdot)$  are non-decreasing univariate functions. While we establish (and explicitly identify) the two functions  $\kappa(n)$  and  $\zeta(n)$  for each of the stable or Poisson white noise processes, we conjecture that (1) gives the correct way to cancel out the diverging terms for any Lévy white noise for

a suitable choice of  $\kappa(n)$  and  $\zeta(n)$ . While the term  $\log(m)$  is reminiscent of the amplitude quantization effect, functions  $\kappa(\cdot)$  and  $\zeta(\cdot)$  quantify the compressibility of a given Lévy process: the higher the growth rate of  $\kappa(n)$ , the less compressible the process. If two processes have the same growth rate of  $\kappa(n)$ , then, the  $\zeta(n)$  with the smaller growth rate is the more compressible process.

Finally, we mention the possibility of other alternatives for quantifying compressability of stochastic processes. The notion of *entropy dimension* is one possible alternative which was introduced in [16] for continuous random variables, and later studied for discrete-domain random processes in [17] in the context of compressed sensing. See also  $\epsilon$ -metric entropy of [18],  $\epsilon$ -entropy of [19], and [20] for a general discussion of defining relative entropy for stochastic processes.

The organization of the paper is as follows. We begin by reviewing the preliminaries, including some of the basic definitions and results regarding differential entropy in Section II. Some of the preliminaries on sparse stochastic processes are given in Appendix A for making the paper self-contained. Next, we present our main contributions in Section III wherein we derive the entropy of stable and impulsive Poisson innovation processes.

To facilitate reading of the paper, we have separated the results from their proofs. Some of the key lemmas for obtaining the final claims are explained in Section IV, while the main body of proofs are postponed to Section V.

## II. PRELIMINARIES

The goal of this paper is to define an entropy measure for certain random processes. Hence, we first review the concept of entropy for random variables. For this purpose, we provide the definition of entropy for three main types of probability distributions. The definition of innovation processes (white Lévy noises) and in particular, the stable and Poisson white noises are given in Appendix A.

All the logarithms in this paper are in base  $e$ , and random variables are denoted by capital letters.

### A. Types of Random variables

The main types of probability distributions considered in this paper are absolutely continuous, absolutely discrete, and discrete-continuous, which are defined below.

**Definition 1** (Absolutely Continuous Random Variables). [21] Let  $\mathcal{B}$  be the Borel  $\sigma$ -field of  $\mathbb{R}$  and let  $X$  be a real-valued random variable that is measurable with respect to  $\mathcal{B}$ . We call  $X$  an absolutely continuous random variable if its probability measure  $\mu$ , induced on  $(\mathbb{R}, \mathcal{B})$ , is absolutely continuous with respect to the Lebesgue measure for  $\mathcal{B}$  (i.e.,  $\mu(\mathcal{A}) = 0$  for all  $\mathcal{A} \in \mathcal{B}$  with zero Lebesgue measure). We denote the set of all absolutely continuous distributions by  $\mathcal{AC}$ .

The Radon-Nikodym theorem implies that for each  $X \in \mathcal{AC}$  there exists a  $\mathcal{B}$ -measurable function  $p : \mathbb{R} \rightarrow [0, \infty)$ , such that for all  $\mathcal{A} \in \mathcal{B}$  we have that

$$\Pr \{X \in \mathcal{A}\} = \int_{\mathcal{A}} p(x) dx.$$

The function  $p$  is called the probability density of  $X$ . The property  $X \in \mathcal{AC}$  is alternatively written as  $p \in \mathcal{AC}$ . [20, p. 21]

**Definition 2** (Absolutely Discrete Random Variable). [21] A random variable  $X$  is called absolutely discrete if it takes values in a countable alphabet set  $\mathcal{X} \subset \mathbb{R}$ .

**Definition 3** (Discrete-Continuous Random Variable). [21] A random variable  $X$  is called discrete-continuous with parameters  $(p_c, P_D, \Pr \{X \in \mathcal{D}\})$  if there exists a countable set  $\mathcal{D}$ , an absolutely discrete probability distribution  $P_D$ , whose support is  $\mathcal{D}$ , and an absolutely continuous distribution  $p_c \in \mathcal{AC}$  such that

$$0 \leq \Pr \{X \in \mathcal{D}\} \leq 1,$$

as well as for every Borel set  $\mathcal{A}$  in  $\mathbb{R}$  we have that

$$\Pr \{X \in \mathcal{A} | X \notin \mathcal{D}\} = \int_{\mathcal{A}} p_c(x) dx, \quad \Pr \{X \in \mathcal{A} | X \in \mathcal{D}\} = \sum_{x \in \mathcal{D} \cap \mathcal{A}} P_D[x].$$

It is clear that we can write the distribution of a discrete-continuous random variable  $X$ ,  $p_X$ , as follows:

$$p_X(x) = \Pr \{X \in \mathcal{D}\} P_d(x) + (1 - \Pr \{X \in \mathcal{D}\}) p_c(x),$$

where  $p_c \in \mathcal{AC}$ , and  $P_d$  is the probability mass function of the discrete part, which is a convex combination of Dirac's delta functions.

In this paper, the probability mass function of absolutely discrete random variables is denoted by capital letters like  $P$  and  $Q$ , while the probability density function of absolutely continuous or discrete-continuous random variables is denoted by lowercase letters like  $p$  and  $q$ .

### B. Definition of Entropy

We first define the entropy and differential entropy for absolutely discrete and absolutely continuous random variables, respectively. Next, we define entropy dimension for discrete-continuous random variables via amplitude quantization.

**Definition 4** (Entropy and Differential Entropy). [22, Chapter 2] We define entropy  $H(X)$ , or  $H(P_X)$  for a discrete random variable  $X$  with distribution  $P_X[x]$  as

$$H(P_X) = H(X) := \sum_x P_X[x] \log \frac{1}{P_X[x]},$$

if the summation converges. For an absolutely continuous random variable  $X$  with distribution  $p_X(x) \in \mathcal{AC}$ , we define differential entropy  $h(X)$ , or  $h(p_X)$  as

$$h(p_X) = h(X) := \int_{\mathbb{R}} p_X(x) \log \frac{1}{p_X(x)} dx,$$

if

$$\int_{\mathbb{R}} p_X(x) \log \left| \frac{1}{p_X(x)} \right| dx < \infty.$$

Next, we identify a class of absolutely continuous probability distributions, and show that differential entropy is uniformly convergent over this space under the total variation distance metric.

**Definition 5.** [23] Given  $\alpha, m, v \in (0, \infty)$ , we define  $(\alpha, v, m)$ - $\mathcal{AC}$  to be the class of all  $p \in \mathcal{AC}$  such that the corresponding density function  $p : \mathbb{R} \mapsto [0, \infty)$  satisfies

$$\int_{\mathbb{R}} |x|^\alpha p(x) dx \leq v,$$

$$\text{ess sup}_{x \in \mathbb{R}} p(x) \leq m.$$

**Theorem 1.** [23] The differential entropy of any distribution in  $(\alpha, v, m)$ - $\mathcal{AC}$  is well-defined, and for all  $p_X, p_Y \in (\alpha, v, m)$ - $\mathcal{AC}$  satisfying  $\|p_X - p_Y\|_1 \leq m$ , we have that

$$|h(p_X) - h(p_Y)| \leq c_1 D_{X,Y} + c_2 D_{X,Y} \log \frac{1}{D_{X,Y}},$$

where

$$D_{X,Y} = \|p_X - p_Y\|_1 := \int_{\mathbb{R}} |p_X(x) - p_Y(x)| dx,$$

and

$$c_1 = \frac{1}{\alpha} |\log(2\alpha v)| + |\log(me)| + \log \frac{e}{2} + \log \left[ 2\Gamma \left( 1 + \frac{1}{\alpha} \right) \right] + \frac{1}{\alpha} + 1,$$

$$c_2 = \frac{1}{\alpha} + 2.$$

Now, we define the quantization of a random variable in amplitude domain.

**Definition 6** (Quantization of Random Variables). *The quantized version of a random variable  $X$  with the step size  $1/m$  ( $m \in \mathbb{Z}$ ) is defined as*

$$[X]_m = \left\lfloor \frac{1}{2} + mX \right\rfloor.$$

Thus,  $[X]_m$  has the distribution  $P_{X;m}$  given by

$$P_{X;m}[i] := \Pr \left\{ [X]_m = \frac{i}{m} \right\} = \int_{\left[ \frac{i-0.5}{m}, \frac{i+0.5}{m} \right)} p_X(x) dx.$$

Also, we define an absolutely continuous random variable  $\tilde{X}_m$  with distribution  $q_{X;m} \in \mathcal{AC}$  as follows

$$q_{X;m}(x) = m P_{X;m}[i], \quad x \in \left[ \frac{i-0.5}{m}, \frac{i+0.5}{m} \right).$$

The following theorem, proved in [16], measures the entropy of quantized discrete-continuous random variables by defining two notions of *entropy dimension* and *dimensional entropy*.

**Theorem 2.** [16] *Let  $X$  be a discrete-continuous random variable defined by the triplet  $(p_c, P_D, \Pr \{X \in \mathcal{D}\})$ . Define absolutely discrete random variable  $X_D \sim P_D$  and absolutely continuous random variable  $X_c \sim p_c$ . If*

$$H([X]_1) < \infty, \quad H(X_D) < \infty, \quad \int_{\mathbb{R}} p_c(x) \left| \log \frac{1}{p_c(x)} \right| dx < \infty,$$

where  $[X]_1$  is the quantized version of  $X$  with step size 1, then the entropy of quantized  $X$  with step size  $1/m$  can be expressed as

$$H([X]_m) = d \log m + d h(X_c) + (1-d) H(X_D) + H_2(d) + o_m(1),$$

where  $o_m(1)$  vanishes as  $m \rightarrow \infty$ ,  $d = \Pr \{X \notin \mathcal{D}\}$  and  $H_2(d) \triangleq d \log 1/d + (1-d) \log [1/(1-d)]$  is the entropy of a discrete random variable with probability distribution  $\{d, 1-d\}$ . The variable  $d$  is called the **Entropy Dimension** of  $X$ .

**Remark 1.** [16] According to Theorem 2, if a random variable  $X$  is absolutely continuous with distribution  $p_c$ , then, the entropy of quantized  $X$  with step size  $1/m$  is

$$H([X]_m) = \log m + h(p_c) + o_m(1),$$

provided that

$$H([X]_1) < \infty, \quad \int_{\mathbb{R}} p_c(x) \left| \log \frac{1}{p_c(x)} \right| dx < \infty,$$

where  $h(X)$  is the differential entropy of  $X$ .

Finally, we point out that in [21], the entropy is defined for a *mixed-pair*  $Z := (X, Y)$ , where  $X$  is an absolutely discrete random variable with distribution  $P$  over the sample space  $\mathcal{X}$  and  $Y$  is an absolutely continuous random variable with distribution  $p$ . The distribution of  $Y$  conditioned to  $X = n$  is denoted by  $p_n(x)$  and it is assumed to be absolutely continuous. The entropy of  $Z$ ,  $H(Z)$ , is defined as

$$H(Z) := H(X) + \sum_{n \in \mathcal{X}} P[n] h(p_n),$$

where  $H(\cdot)$  and  $h(\cdot)$  are the entropy and the differential entropy, respectively.

### III. MAIN RESULTS

We first propose a criterion for the compressibility of continuous random processes, and study its operational meaning from the viewpoint of source coding. Then, in Section III-B, we evaluate the compressibility criterion for stable and Poisson innovation processes. Finally, we present a qualitative comparison between the compressibility of the considered innovation processes in Section III-C.

#### A. Compressibility via quantization

A generic continuous-domain stochastic process is spread over the continuum of time and amplitude. Hence, we doubly discretize the process by applying time and amplitude quantization. This enables us to utilize the conventional definition of the entropy measure. Then, we monitor the entropy trends as the quantizations become finer.

The amplitude quantization was previously defined in Definition 6. The time quantization, or equivalently, the sampling in time shall be defined in a similar fashion:



**Definition 7** (Time Quantization). *The time quantization with step size  $1/n$  of a white Lévy noise (an innovation process)  $X(t)$  is defined as the sequence  $\{X_i^{(n)}\}_{i \in \mathbb{Z}}$  of random variables*

$$X_i^{(n)} = \langle X, \phi_{i,n} \rangle,$$

where  $\phi_{i,n}(t) = \phi(nt - i + 1)$  and

$$\phi(t) = \begin{cases} 1 & t \in [0, 1) \\ 0 & t \notin [0, 1) \end{cases}. \quad (2)$$

**Remark 2.** *Observe that  $\phi(t)$  is not a member of  $\mathcal{S}(\mathbb{R})$  as defined in Definition 8. Hence, strictly speaking, for a white Lévy noise  $X(t)$  with sample space  $\mathcal{S}'(\mathbb{R})$ , we cannot automatically define  $\langle X, \phi \rangle$  based on Definition 10. However, the random variables  $X_i^{(n)}$  could be easily interpreted as the increments of the Lévy process corresponding to this white noise. Alternatively, one can define  $\langle X, \phi \rangle$  as the limit of  $\langle X, \psi_n \rangle$  as  $n \rightarrow \infty$ , when  $\{\psi_n(t)\}_{n=1}^\infty \subset \mathcal{S}(\mathbb{R})$  satisfy  $\lim_{n \rightarrow \infty} \int_{\mathbb{R}} |\psi_n(t) - \phi(t)| dt = 0$ . The existence of the limit  $\langle X, \psi_n \rangle$  when  $X(t)$  is either a stable or a Poisson innovation process, is trivial.*

Our next step, is to find the entropy rate of a (doubly) quantized random process. Let  $X(t)$  be a white Lévy noise and define the random variables

$$\tilde{X}_i^{m,n} := \left( \left[ X_{n(i-1)+1}^{(n)} \right]_m, \dots, \left[ X_{ni}^{(n)} \right]_m \right),$$

where  $X_i^{(n)}$  refers to the time quantization of the process (Definition 7) followed by amplitude quantization in the form  $\left[ X_i^{(n)} \right]_m$  as shown in Definition 6. The time quantization in  $\tilde{X}_i^{m,n}$  spans the interval  $t \in [i-1, i)$  of the process  $X(t)$  via  $n$  random variables  $\left[ X_{n(i-1)+j}^{(n)} \right]_m$ ,  $j = 1, \dots, n$ . Thus, the sequence  $\{\tilde{X}_i^{m,n}\}_{i \in \mathbb{Z}}$  represents the innovation process over the whole real axis  $t \in \mathbb{R}$  in a quantized way. We evaluate the entropy rate (entropy per unit interval of time) for this source sequence via

$$\mathcal{H}_{m,n}(X) := \lim_{T \rightarrow \infty} \frac{H(\tilde{X}_{-T+1}^{m,n}, \dots, \tilde{X}_T^{m,n})}{2T}, \quad (3)$$

where  $H(\cdot)$  stands for the discrete entropy. The above definition has an operational meaning in terms of the number of bits required for asymptotic lossless compression of the source as  $T$  tends to  $\infty$ . Since  $\left[ X_i^{(n)} \right]_m$ 's depend on equi-length and non-overlapping time intervals of the white noise, they are independent and identically distributed (Lemma 5). Therefore,

$$\mathcal{H}_{m,n}(X) = H(\tilde{X}_1^{m,n}) = nH\left(\left[ X_1^{(n)} \right]_m\right). \quad (4)$$

To compensate for the quantization effect, we shall study the behavior of  $\mathcal{H}_{m,n}(X)$  as  $m, n \rightarrow \infty$ .

### B. Compressibility of stable and Poisson white noises

The following two theorems demonstrate the asymptotic entropy rate  $\mathcal{H}_{m,n}(X)$  of the quantized white noise under stable and Poisson distributions, when the quantization steps vanish.

**Theorem 3.** *Let  $X(t)$  be a stable white noise with parameters  $(\alpha, \beta, \sigma, \mu)$ . We define*

$$X_0 := \langle X, \phi \rangle = \int_0^1 X(t) dt,$$

where  $\phi$  is the function defined in (2). Then,

$$\lim_{m,n, \frac{m}{\sqrt[n]{n}} \rightarrow \infty} \left( \frac{\mathcal{H}_{m,n}(X)}{n} - \log \frac{m}{\sqrt[n]{n}} \right) = h(X_0),$$

where  $\mathcal{H}_{m,n}(X)$  is defined in (4), and  $h(X_0)$  is the differential entropy of continuous random variable  $X_0$ .

The proof is given in Section V-A.

**Theorem 4.** *Let  $X(t)$  be a Poisson white noise with rate  $\lambda$  and amplitude distribution  $p_A$  such that:*

$$p_A \in (\alpha, M, V)\text{-}\mathcal{AC}, \quad (5)$$

for some positive constants  $\alpha, M, V$  and

$$\int_{\mathbb{R}} p_A(x) \left| \log \frac{1}{p_A(x)} \right| dx < \infty, \quad H([A]_1) < \infty,$$

where  $A \sim p_A$  and  $[A]_1$  is the quantization of  $A$  with step size one. Then,

$$\lim_{\substack{m,n \rightarrow \infty \\ \frac{\log m}{n}, \frac{\log n}{m} \rightarrow 0}} \left( \frac{\mathcal{H}_{m,n}(X)}{\lambda} - \log(mn) \right) = h(A) - \log \lambda + 1.$$

The proof is given in Section V-B.

### C. Comparison and discussion

Both of our results in Theorems 3 and 4 can be stated as the limits of the form

$$\frac{\mathcal{H}_{m,n}(X)}{\kappa(n)} - \log(m) - \zeta(n), \quad (6)$$

where  $\kappa(\cdot)$  and  $\zeta(\cdot)$  are non-decreasing univariate functions. Essentially, the limits are evaluated for  $m, n \rightarrow \infty$ , while  $m$  (representing the fineness in amplitude quantization) and  $n$  (representing the fineness in time quantization) cannot independently vary. To better represent the constraints

on  $(m, n)$  pairs in each of our results, let us define the set of all Feasible Sequence of Pairs (FSP) as

$$\text{FSP}_{\text{Stable}}^\alpha = \left\{ \left\{ (m_k, n_k) \right\}_{k=1}^\infty \mid n_k, \frac{m_k}{\sqrt[n_k]{n_k}} \xrightarrow{k \rightarrow \infty} \infty \right\} \quad (7)$$

and

$$\text{FSP}_{\text{Poisson}} = \left\{ \left\{ (m_k, n_k) \right\}_{k=1}^\infty \mid \frac{m_k}{1+\log n_k}, \frac{n_k}{1+\log m_k} \xrightarrow{k \rightarrow \infty} \infty \right\}. \quad (8)$$

In order to compare the compressibility of two white noise processes  $X_1(t)$  and  $X_2(t)$  corresponding to  $\kappa_i(\cdot)$  and  $\zeta_i(\cdot)$  for  $i = 1, 2$ , we compare the asymptotic behavior of their  $\mathcal{H}_{m,n}(X)$ . For this purpose, we first have to make sure that the used sequence of  $(m, n)$  is feasible for both processes. Next, we check  $\lim_{n \rightarrow \infty} \frac{\kappa_1(n)}{\kappa_2(n)}$ ; if the latter is 0 (similarly,  $\infty$ ), we conclude that  $X_1$  (similarly,  $X_2$ ) is more compressible. If this limit is a finite and non-zero number like  $\tilde{\kappa}$ , then, we need to examine  $\lim_{n \rightarrow \infty} \zeta_2(n) - \tilde{\kappa} \zeta_1(n)$ ; if this limit approaches  $\infty$  (similarly,  $-\infty$ ), we again conclude that  $X_1$  (similarly,  $X_2$ ) is more compressible. If neither of these tests is decisive, we have to check the limiting values of (6) for each process. The following theorem summarizes the comparisons within the family of stable and Poisson white noises.

**Theorem 5.** *The following statements apply to Poisson and stable white noises.*

- i) *A stable white noise becomes less compressible as its stability parameter  $\alpha$  increases. Let  $X_1$  and  $X_2$  be stable white noises with stability parameters  $0 < \alpha_1 < \alpha_2 \leq 2$  and let  $0 < \bar{\alpha} \leq \alpha_1$ ; then, for all  $\{(m_k, n_k)\}_{k=1}^\infty \in \text{FSP}_{\text{Stable}}^{\bar{\alpha}}$  we have that*

$$\lim_{k \rightarrow \infty} \left( \mathcal{H}_{m_k, n_k}(X_2) - \mathcal{H}_{m_k, n_k}(X_1) \right) = +\infty. \quad (9)$$

- ii) *A Poisson white noise becomes less compressible as its rate parameter  $\lambda$  increases. Let  $X_1$  and  $X_2$  be Poisson white noises with rate parameters  $0 < \lambda_1 < \lambda_2$ ; then, for all  $\{(m_k, n_k)\}_{k=1}^\infty \in \text{FSP}_{\text{Poisson}}$  we have that*

$$\lim_{k \rightarrow \infty} \left( \mathcal{H}_{m_k, n_k}(X_2) - \mathcal{H}_{m_k, n_k}(X_1) \right) = +\infty. \quad (10)$$

- iii) *The Poisson white noises are always more compressible than stable white noises. Let  $X$  be an arbitrary Poisson white noise and let  $Y$  be a stable white noise with stability parameter  $\alpha \in (0, 2]$ ; then, for all*

$$\{(m_k, n_k)\}_{k=1}^\infty \in \text{FSP}_{\text{Poisson}} \cap \text{FSP}_{\text{Stable}}^\alpha$$

we have that

$$\lim_{k \rightarrow \infty} \frac{\mathcal{H}_{m_k, n_k}(X)}{\mathcal{H}_{m_k, n_k}(Y)} = 0. \quad (11)$$

The proof can be found in Section V-C.

We should mention that Parts i) and ii) of Theorem 5 confirm with previously studied notions of compressibility in [7], [9], [10]. However, Part iii) deviates from the available literature by identifying the Poisson white noises as more compressible than heavy-tailed stable white noises. This difference is fundamental and is caused by basing the compressibility definition on the probability concentration properties of processes rather than their amplitude distribution.

#### IV. SOME USEFUL LEMMAS

In Section IV-A, we provide a useful theorem for Poisson white noises, and in Section IV-B, we give some lemmas about quantization of a random variable in amplitude domain. These results are used in proof of the main results.

##### A. Poisson White Noise

The following theorem states a feature of integrated Poisson white noise in a small interval.

**Theorem 6.** *Let  $X(t)$  be a Poisson white noise with rate  $\lambda$  and amplitude distribution  $p_A \in \mathcal{AC}$ . If  $Y_n$  is defined as*

$$Y_n = \langle X, \phi(nt) \rangle = \int_0^{\frac{1}{n}} X(t) dt, \quad (12)$$

*then the probability distribution of  $Y_n$  is*

$$p_{Y_n}(x) = e^{-\frac{\lambda}{n}} \delta(x) + \left(1 - e^{-\frac{\lambda}{n}}\right) p_{A_n}(x), \quad (13)$$

*where  $A_n$  is a random variable with probability distribution  $p_{A_n}$  such that*

$$p_{A_n}(x) = \frac{1}{e^{\frac{\lambda}{n}} - 1} \sum_{k=1}^{\infty} \frac{\left(\frac{\lambda}{n}\right)^k}{k!} (\overbrace{p_A * \dots * p_A}^k)(x), \quad (14)$$

$$\int_{\mathbb{R}} |p_{A_n}(x) - p_A(x)| dx \leq 2 \frac{e^{\frac{\lambda}{n}} - \frac{\lambda}{n} - 1}{e^{\frac{\lambda}{n}} - 1}, \quad (15)$$

$$\lim_{n \rightarrow \infty} \mathbb{E}[|A_n|^\alpha] = \mathbb{E}[|A|^\alpha], \quad \forall \alpha \in (0, \infty), \quad (16)$$

*where  $*$  is the convolution operator. In addition, in (15), the upper bound on total variation distance vanishes as  $n$  tends to infinity.*

The proof is given in Section V-D.

### B. Amplitude Quantization

In the following lemma, we show that the total variation distance between two variables decreases by quantizing; furthermore, moments of a quantized random variable tends to the moments of the original random variable, as the quantization step size vanishes.

**Lemma 1.** *Let random variables  $X \sim p_X$  and  $Y \sim p_Y$  be absolutely continuous, and  $[X]_m \sim P_{X;m}$  and  $[Y]_m \sim P_{Y;m}$  be their quantized version, defined in Definition 6, respectively. Then for all  $m \in (0, \infty)$  we have*

$$\int_{\mathbb{R}} |q_{X;m}(x) - q_{Y;m}(x)| \, dx \leq \int_{\mathbb{R}} |p_X(x) - p_Y(x)| \, dx, \quad (17)$$

where  $\tilde{X}_m \sim q_{X;m}$  and  $\tilde{Y}_m \sim q_{Y;m}$  are random variables defined in Definition 6. In addition, for any  $\alpha \in (0, \infty)$ , and  $m \geq 4$  we have

$$\mathbb{E} \left[ \left| \tilde{X}_m \right|^\alpha \right] \leq \left( \frac{2}{\sqrt{m}} \right)^\alpha + e^{\frac{\alpha}{\sqrt{m}}} \mathbb{E} [|X|^\alpha], \quad (18)$$

$$\Pr \left\{ |X| > \frac{1}{\sqrt{m}} \right\} e^{-2\frac{\alpha}{\sqrt{m}}} \mathbb{E} [|X|^\alpha] \leq \mathbb{E} \left[ \left| \tilde{X}_m \right|^\alpha \right], \quad (19)$$

provided that  $\mathbb{E} [|X|^\alpha]$  exists.

The proof is given in Section V-E.

**Corollary 1.** *Let random variable  $X \sim p_X$  be absolutely continuous random variable, and  $\tilde{X}_m \sim q_{X;m}$  be the random variable defined in Definition 6. Then, we have that*

$$\lim_{m \rightarrow \infty} \mathbb{E} \left[ \left| \tilde{X}_m \right|^\alpha \right] = \mathbb{E} [|X|^\alpha],$$

if  $\mathbb{E} [|X|^\alpha]$  exists.

*Proof.* Since,  $p_X \in \mathcal{AC}$ , we conclude that  $\Pr \{|X| \leq 1/\sqrt{m}\}$  vanishes as  $m$  tends to  $\infty$ . Hence, the corollary achieved from Lemma 1.  $\square$

The following lemma discusses the entropy of quantized version of multiplies of an absolutely continuous random variable.

**Lemma 2.** *Let  $X \sim p$  be an absolutely continuous random variable, and  $m \in (0, \infty)$  be arbitrary. If  $H([X]_m)$  exists, then for all  $a \in (0, \infty)$ , we have*

$$H \left( [aX]_{\frac{m}{a}} \right) = H([X]_m).$$

The lemma is proved in Section V-F.

In the following lemma, we extend Remark 1 to an arbitrary shifted absolutely continuous random variables.

**Lemma 3.** *Let  $X \sim p$  be an absolutely continuous random variable with a piecewise continuous density function  $p(x)$ . For an arbitrary sequence  $\{c_m\}_{m=1}^\infty \subset \mathbb{R}$ , we have*

$$\lim_{m \rightarrow \infty} H([X + c_m]_m) - \log m = h(X),$$

*provided that*

$$H([X]_1) < \infty, \quad \int_{\mathbb{R}} p(x) \left| \log \frac{1}{p(x)} \right| dx < \infty, \quad \text{and} \quad \text{ess sup}_{x \in \mathbb{R}} p(x) = L < \infty,$$

*where  $[X]_1$  is the quantized version of  $X$  with step size 1.*

The proof can be found in Section V-G.

## V. PROOFS

### A. Proof of Theorem 3

In order to prove the theorem, we need the following lemma, which is proved in Appendix B.

**Lemma 4.** *Let  $X$  be a stable white noise with parameters  $(\alpha, \beta, \sigma, \mu)$ . Define random variable  $X_0$  with distribution  $p_{X_0}$  as follows:*

$$X_0 = \langle X, \phi \rangle = \int_0^1 X(t) dt,$$

*where  $\phi$  is the function defined in (2). Then,*

$$p_{X_0} \in \mathcal{AC}, \tag{20}$$

$$\text{ess sup}_{x \in \mathbb{R}} p_{X_0}(x) = L < \infty, \tag{21}$$

$$p_{X_0}(x) \text{ is a piecewise continuous function,} \tag{22}$$

$$\int_{\mathbb{R}} p_{X_0}(x) \left| \log \frac{1}{p_{X_0}(x)} \right| dx < \infty, \tag{23}$$

$$H([X_0]_1) < \infty. \tag{24}$$

Now we can prove Theorem 3.

*Proof.* Using Definition 12, the characteristic function of  $X_1^{(n)}$  is as follows:

$$\widehat{p}_{X_1^{(n)}}(\omega) = \begin{cases} e^{-\sigma^\alpha |\omega|^\alpha \left(\frac{1}{n} - j\frac{1}{n}\beta \operatorname{sgn}(\omega) \tan \frac{\pi\alpha}{2}\right) + j\mu \frac{\omega}{n}} & \alpha \neq 1 \\ e^{-\sigma |\omega| \left(\frac{1}{n} + j\frac{1}{n}\beta \frac{2}{\pi} \operatorname{sgn}(\omega) \ln |\omega|\right) + j\mu \frac{\omega}{n}} & \alpha = 1 \end{cases}.$$

From the definition of  $X_0$ , we can write

$$\widehat{p}_{X_0}(\omega) = \begin{cases} e^{-\sigma^\alpha |\omega|^\alpha \left(1 - j\beta \operatorname{sgn}(\omega) \tan \frac{\pi\alpha}{2}\right) + j\mu \omega} & \alpha \neq 1 \\ e^{-\sigma |\omega| \left(1 + j\beta \frac{2}{\pi} \operatorname{sgn}(\omega) \ln |\omega|\right) + j\mu \omega} & \alpha = 1 \end{cases}.$$

Thus, we obtain that

$$\widehat{p}_{X_1^{(n)}}(\omega) = \widehat{p}_{X_0}\left(\frac{\omega}{\sqrt[n]{n}}\right) e^{-j\omega c_n}$$

where

$$c_n = \begin{cases} \mu \left(\frac{1}{\sqrt[n]{n}} - \frac{1}{n}\right) & \alpha \neq 1 \\ \frac{2}{\pi} \sigma \beta \frac{\ln n}{n} & \alpha = 1 \end{cases},$$

Therefore,  $X_1^{(n)}$  can be written with respect to  $X_0$  as follows:

$$X_1^{(n)} = \frac{X_0 - b_n}{\sqrt[n]{n}} \quad (\text{in distribution}),$$

where  $b_n = c_n \sqrt[n]{n}$ . Hence,

$$\mathbb{H}\left(\left[X_1^{(n)}\right]_m\right) = \mathbb{H}\left(\left[\frac{X_0 - b_n}{\sqrt[n]{n}}\right]_m\right) \quad (25)$$

Therefore, from Lemma 2 we obtain that

$$\mathbb{H}\left(\left[X_1^{(n)}\right]_m\right) = \mathbb{H}\left(\left[X_0 - b_n\right]_{\frac{m}{\sqrt[n]{n}}}\right). \quad (26)$$

Lemma 4 shows that the distribution of  $X_0$  satisfies the properties of Lemma 3. Thus, from Lemma 3 we obtain that when  $\frac{m}{\sqrt[n]{n}}$  tend to  $\infty$ , we have

$$\lim_{\frac{m}{\sqrt[n]{n}} \rightarrow \infty} \mathbb{H}\left(\left[X_0 - b_n\right]_{\frac{m}{\sqrt[n]{n}}}\right) - \log \frac{m}{\sqrt[n]{n}} = h(X_0). \quad (27)$$

Statement of the theorem follows from (25)-(27).  $\square$

#### B. Proof of Theorem 4

We define the discrete distribution  $P_{X_1^{(n)};m}$  as in Definition 6. Therefore,

$$\begin{aligned} n\mathbb{H}\left(\left[X_1^{(n)}\right]_m\right) &= n \sum_{i \in \mathbb{Z}} P_{X_1^{(n)};m}[i] \log \frac{1}{P_{X_1^{(n)};m}[i]} \\ &= n P_{X_1^{(n)};m}[0] \log \frac{1}{P_{X_1^{(n)};m}[0]} + n \sum_{i \in \mathbb{Z} \setminus \{0\}} P_{X_1^{(n)};m}[i] \log \frac{1}{P_{X_1^{(n)};m}[i]}. \end{aligned}$$

We are going to prove that

$$\lim_{m,n \rightarrow \infty} n P_{X_1^{(n)};m}[0] \log \frac{1}{P_{X_1^{(n)};m}[0]} = \lambda, \quad (28)$$

$$\lim_{\substack{m,n \rightarrow \infty \\ \frac{\log n}{m}, \frac{\log m}{n} \rightarrow 0}} n \sum_{i \in \mathbb{Z} \setminus \{0\}} P_{X_1^{(n)};m}[i] \log \frac{1}{P_{X_1^{(n)};m}[i]} - \lambda \log mn = \lambda h(A) - \lambda \log \lambda, \quad (29)$$

where  $h(A)$  is the differential entropy of random variable  $A$ . Proving the above two equalities, will imply the statement of the theorem.

*Proof of (28):* From Theorem 6, we obtain that

$$P_{X_1^{(n)};m}[0] = e^{-\frac{\lambda}{n}} + (1 - e^{-\frac{\lambda}{n}}) P_{A_n;m}[0], \quad (30)$$

for some random variable  $A_n$  with features mentioned in (14)-(16). For every  $m$  we have

$$\int_{-\frac{1}{2m}}^{\frac{1}{2m}} |p_{A_n}(x) - p_A(x)| dx \leq \int_{\mathbb{R}} |p_{A_n}(x) - p_A(x)| dx.$$

Hence,

$$\int_{-\frac{1}{2m}}^{\frac{1}{2m}} p_{A_n}(x) dx \leq \int_{-\frac{1}{2m}}^{\frac{1}{2m}} p_A(x) dx + \int_{\mathbb{R}} |p_{A_n}(x) - p_A(x)| dx \leq \frac{M}{m} + 2 \frac{e^{\frac{\lambda}{n}} - \frac{\lambda}{n} - 1}{e^{\frac{\lambda}{n}} - 1}, \quad (31)$$

where (31) is true due to (15), and the fact that  $p_A(x) \leq M$  almost everywhere. Thus, we obtain that

$$\lim_{m,n \rightarrow \infty} \int_{-\frac{1}{2m}}^{\frac{1}{2m}} p_{A_n}(x) dx = 0.$$

Therefore, from the definition of  $P_{A_n;m}[0]$  defined in Definition 6, we obtain that

$$\lim_{m,n \rightarrow \infty} P_{A_n;m}[0] = 0. \quad (32)$$

Therefore, from (30), we can write that

$$\lim_{m,n \rightarrow \infty} P_{X_1^{(n)};m}[0] = 1. \quad (33)$$

Hence, if we prove that

$$\lim_{m,n \rightarrow \infty} n \log \frac{1}{P_{X_1^{(n)};m}[0]} = \lambda, \quad (34)$$

we can conclude from (33) that

$$\lim_{m,n \rightarrow \infty} n P_{X_1^{(n)};m}[0] \log \frac{1}{P_{X_1^{(n)};m}[0]} = \lim_{m,n \rightarrow \infty} n \log \frac{1}{P_{X_1^{(n)};m}[0]} = \lambda.$$



This will complete the proof of (28). Thus, it only remains to prove (34). Again, by substituting the value of  $P_{X_1;m}[0]$  from (30), we have

$$\begin{aligned} \lim_{m,n \rightarrow \infty} n \log \frac{1}{P_{X_1^{(n)};m}[0]} &= - \lim_{m,n \rightarrow \infty} n \log \left( e^{-\frac{\lambda}{n}} + (1 - e^{-\frac{\lambda}{n}}) P_{A_n;m}[0] \right) \\ &= - \lim_{m,n \rightarrow \infty} n \log \left( 1 + (e^{-\frac{\lambda}{n}} - 1) (1 - P_{A_n;m}[0]) \right) \\ &= - \lim_{m,n \rightarrow \infty} n (e^{-\frac{\lambda}{n}} - 1) (1 - P_{A_n;m}[0]) \end{aligned} \quad (35)$$

$$= - \lim_{m,n \rightarrow \infty} n (e^{-\frac{\lambda}{n}} - 1) = \lambda. \quad (36)$$

where (35) follows from Taylor series of  $\log(1+x)$  near 0 for  $x = (e^{-\lambda/n} - 1)(1 - P_{A_n;m}[0])$ ; observe that  $e^{-\lambda/n} - 1$  is close to zero for large values of  $n$ , and (36) is obtained from (32).

*Proof of (29):* From Theorem 6 we obtain that

$$P_{X_1^{(n)};m}[i] = \left(1 - e^{-\frac{\lambda}{n}}\right) P_{A_n;m}[i], \quad \forall i \in \mathbb{Z} \setminus \{0\}.$$

By substituting the value of  $P_{X_1^{(n)};m}[i]$  in terms of  $P_{A_n;m}[i]$ , we have

$$\begin{aligned} n \sum_{i \in \mathbb{Z} \setminus \{0\}} P_{X_1^{(n)};m}[i] \log \frac{1}{P_{X_1^{(n)};m}[i]} &= n \left(1 - e^{-\frac{\lambda}{n}}\right) (1 - P_{A_n;m}[0]) \log \frac{1}{1 - e^{-\frac{\lambda}{n}}} \\ &\quad - n \left(1 - e^{-\frac{\lambda}{n}}\right) P_{A_n;m}[0] \log \frac{1}{P_{A_n;m}[0]} \\ &\quad + n \left(1 - e^{-\frac{\lambda}{n}}\right) \sum_{i \in \mathbb{Z}} P_{A_n;m}[i] \log \frac{1}{P_{A_n;m}[i]}. \end{aligned}$$

Therefore, in order to prove (29), it suffices to show that

$$\lim_{m,n \rightarrow \infty} n \left(1 - e^{-\frac{\lambda}{n}}\right) P_{A_n;m}[0] \log P_{A_n;m}[0] = 0, \quad (37)$$

$$\lim_{m,n, \frac{m}{\log n} \rightarrow \infty} n \left(1 - e^{-\frac{\lambda}{n}}\right) (1 - P_{A_n;m}[0]) \log \frac{1}{1 - e^{-\frac{\lambda}{n}}} - \lambda \log n = -\lambda \log \lambda, \quad (38)$$

$$\lim_{m,n, \frac{n}{\log m} \rightarrow \infty} n \left(1 - e^{-\frac{\lambda}{n}}\right) \sum_{i \in \mathbb{Z}} P_{A_n;m}[i] \log \frac{1}{P_{A_n;m}[i]} - \lambda \log m = \lambda h(A). \quad (39)$$

*Proof of (37):* It is obtained from (32), and the fact that

$$\lim_{n \rightarrow \infty} n \left(1 - e^{-\frac{\lambda}{n}}\right) = \lambda. \quad (40)$$

*Proof of (38):* Let us add and subtract the term  $n(1 - e^{-\lambda/n})(1 - P_{A_n;m}[0]) \log n$  to the left hand side of (38) as follows:

$$\begin{aligned} n(1 - e^{-\frac{\lambda}{n}})(1 - P_{A_n;m}[0]) \log \frac{1}{1 - e^{-\frac{\lambda}{n}}} - \lambda \log n &= n(1 - e^{-\frac{\lambda}{n}})(1 - P_{A_n;m}[0]) \log \frac{1}{n(1 - e^{-\frac{\lambda}{n}})} \\ &\quad + [n(1 - e^{-\frac{\lambda}{n}})(1 - P_{A_n;m}[0]) - \lambda] \log n \end{aligned}$$

From (32), and (40), we can write

$$\lim_{m, n \rightarrow \infty} n(1 - e^{-\frac{\lambda}{n}})(1 - P_{A_n;m}[0]) \log \frac{1}{n(1 - e^{-\frac{\lambda}{n}})} = -\lambda \log \lambda.$$

Hence, we only need to show that

$$\lim_{\substack{m, n \rightarrow \infty \\ \frac{\log n}{m} \rightarrow 0}} [n(1 - e^{-\frac{\lambda}{n}})(1 - P_{A_n;m}[0]) - \lambda] \log n = 0. \quad (41)$$

We can write the left hand side of (41) as follows:

$$[n(1 - e^{-\frac{\lambda}{n}})(1 - P_{A_n;m}[0]) - \lambda] \log n = [n(1 - e^{-\frac{\lambda}{n}}) - \lambda] \log n \quad (42)$$

$$- n(1 - e^{-\frac{\lambda}{n}}) P_{A_n;m}[0] \log n. \quad (43)$$

For (42), it is obtained that

$$\lim_{n \rightarrow \infty} [n(1 - e^{-\frac{\lambda}{n}}) - \lambda] \log n = 0.$$

For (43), from (31), and the fact that  $0 \leq 1 - e^{-\lambda/n} \leq \lambda/n$ , we can write that

$$0 \leq n(1 - e^{-\frac{\lambda}{n}}) P_{A_n;m}[0] \log n \leq \lambda \left( \frac{M}{m} + 2 \frac{e^{\frac{\lambda}{n}} - \frac{\lambda}{n} - 1}{e^{\frac{\lambda}{n}} - 1} \right) \log n.$$

Thus, (41) is proved due that  $(\log n)/m$  tends to 0, and the fact that

$$\lim_{n \rightarrow \infty} \frac{e^{\frac{\lambda}{n}} - \frac{\lambda}{n} - 1}{e^{\frac{\lambda}{n}} - 1} \log n = 0.$$

Thus, (38) is proved

*Proof of (39):* Let us add and subtract the term  $n(1 - e^{-\lambda/n}) \log m$  to the left hand side of (39) as follows:

$$\begin{aligned} n(1 - e^{-\frac{\lambda}{n}}) \sum_{i \in \mathbb{Z}} P_{A_n;m}[i] \log \frac{1}{P_{A_n;m}[i]} - \lambda \log m \\ = n(1 - e^{-\frac{\lambda}{n}}) \left[ \sum_{i \in \mathbb{Z}} P_{A_n;m}[i] \log \frac{1}{P_{A_n;m}[i]} - \log m \right] + [n(1 - e^{-\frac{\lambda}{n}}) - \lambda] \log m. \end{aligned} \quad (44)$$

From the fact that  $(\log m)/n$  vanishes as  $m, n$  tend to  $\infty$ , we obtain that

$$\lim_{m, n, \frac{n}{\log m} \rightarrow \infty} \left[ n \left( 1 - e^{-\frac{\lambda}{n}} \right) - \lambda \right] \log m = 0. \quad (45)$$

Hence, in order to prove (39), from (40), (44), and (45), it suffices to show that

$$\lim_{m, n \rightarrow \infty} \sum_{i \in \mathbb{Z}} P_{A_n; m}[i] \log \frac{1}{P_{A_n; m}[i]} - \log m = h(A). \quad (46)$$

Remember that in Definition 6, for every arbitrary random variable  $X$ , a random variable  $\tilde{X}_m$  with an absolutely continuous distribution  $q_{X; m}(x)$  was defined. Thus, corresponding to  $A_n$ , we can define random variables  $\tilde{A}_{nm}$  with an absolutely continuous random variable with distribution  $q_{A_n; m}(x)$ . Observe that

$$\begin{aligned} \sum_{i \in \mathbb{Z}} P_{A_n; m}[i] \log \frac{1}{P_{A_n; m}[i]} &= \sum_{i \in \mathbb{Z}} \frac{1}{m} q_{A_n; m}\left(\frac{i}{m}\right) \log \frac{m}{q_{A_n; m}\left(\frac{i}{m}\right)} \\ &= \int_{\mathbb{R}} q_{A_n; m}(x) \log \frac{m}{q_{A_n; m}(x)} dx \\ &= \log m + \int_{\mathbb{R}} q_{A_n; m}(x) \log \frac{1}{q_{A_n; m}(x)} dx \\ &= \log m + h\left(\tilde{A}_{nm}\right), \end{aligned} \quad (47)$$

where  $\tilde{A}_{nm}$  is an absolutely continuous random variable with distribution  $q_{A_n; m}(x)$ . Similarly, for random variable  $A$ , using Definition 6, we can define random variable  $\tilde{A}_m$  with an absolutely continuous distribution  $q_{A; m}(x)$ . Again, we can prove that

$$H([A]_m) := \sum_{i \in \mathbb{Z}} P_{A; m}[i] \log \frac{1}{P_{A; m}[i]} = \log m + h\left(\tilde{A}_m\right). \quad (48)$$

From (48) and Remark 1, we obtain

$$h(A) = \lim_{m \rightarrow \infty} h\left(\tilde{A}_m\right) \quad (49)$$

Hence, from (47) and (49), we obtain that in order to prove (46), it suffices to show that

$$\lim_{m, n \rightarrow \infty} h\left(\tilde{A}_{nm}\right) = \lim_{m \rightarrow \infty} h\left(\tilde{A}_m\right). \quad (50)$$

To show this, we utilize Theorem 1. This theorem reduces convergence in entropy to convergence in total variation distance for a restricted class of distributions. In other words, to show (50), it suffices to show

$$\lim_{m, n \rightarrow \infty} \int_{\mathbb{R}} |q_{A_n; m}(x) - q_{A; m}(x)| dx = 0, \quad (51)$$

as long as we can show that  $q_{A_n;m}(x)$  and  $q_{A;m}(x)$  for all  $m, n \in \mathbb{N}$  belong to the class of distributions given in Definition 5. Remember that in the statement of the theorem, in equation (5), we had assumed that

$$p_A \in (\alpha, M, V)\text{-}\mathcal{AC}, \quad (52)$$

for some positive values for  $\alpha, M, V$ . We show that for all  $m, n \in \mathbb{N}$

$$q_{A_n;m}(x), q_{A;m}(x) \in (\alpha, M, V')\text{-}\mathcal{AC},$$

where  $V' = 2V + 3$ . In other words, for all  $m, n \in \mathbb{N}$  we have

$$q_{A_n;m}(x), q_{A;m}(x) \in \mathcal{AC}, \quad (53)$$

$$q_{A_n;m}(x), q_{A;m}(x) \leq M, \quad (\text{a.e.}), \quad (54)$$

$$\mathbb{E} \left[ \left| \tilde{A}_{nm} \right|^\alpha \right], \mathbb{E} \left[ \left| \tilde{A}_m \right|^\alpha \right] \leq 2V + 3. \quad (55)$$

As a result, it remains to show (51), (53), (54) and (55).

*Proof of (51):* From (17) in Lemma 1, for all  $m, n \in \mathbb{N}$  we have

$$\int_{\mathbb{R}} |q_{A_n;m}(x) - q_{A;m}(x)| \, dx \leq \int_{\mathbb{R}} |p_{A_n}(x) - p_A(x)| \, dx.$$

Thus,

$$\lim_{m,n \rightarrow \infty} \int_{\mathbb{R}} |q_{A_n;m}(x) - q_{A;m}(x)| \, dx \leq \lim_{n \rightarrow \infty} \int_{\mathbb{R}} |p_{A_n}(x) - p_A(x)| \, dx = 0. \quad (56)$$

where (56) follows from Theorem 6. Hence, (51) is proved.

*Proof of (53):* Since the distributions  $q_{A_n;m}$  and  $q_{A;m}$  are combination of step functions, hence, they are absolutely continuous.

*Proof of (54):* Since there exists some  $M$  such that  $p_A(x) \leq M$  for almost all  $x \in \mathbb{R}$ , we obtain that

$$q_{A;m}(x) = mP_{A;m}[i] = m \int_{\frac{i-\frac{1}{2}}{m}}^{\frac{i+\frac{1}{2}}{m}} p_A(x) \, dx \leq m \int_{\frac{i-\frac{1}{2}}{m}}^{\frac{i+\frac{1}{2}}{m}} M \, dx = M.$$

It remains to show that  $q_{A_n;m}(x)$  is bounded too. Similar to the above equation, it suffices to show that  $p_{A_n}(x) \leq M$ . From (14) in Theorem 6, we have that

$$\begin{aligned} p_{A_n}(x) &= \frac{e^{-\frac{\lambda}{n}}}{1 - e^{-\frac{\lambda}{n}}} \sum_{k=1}^{\infty} \frac{\left(\frac{\lambda}{n}\right)^k}{k!} \left( \overbrace{p_A * \cdots * p_A}^k \right)(x) \\ &\leq \frac{e^{-\frac{\lambda}{n}}}{1 - e^{-\frac{\lambda}{n}}} \sum_{k=1}^{\infty} \frac{\left(\frac{\lambda}{n}\right)^k}{k!} M = M, \end{aligned} \quad (57)$$

where (57) follows from the fact that  $p_A$  is bounded, and for any arbitrary absolutely continuous random variables  $T, S$  with distributions  $p_T, p_S$ , respectively, we have  $(p_T * p_S)(x) \leq M$  for all  $x \in \mathbb{R}$  if  $p_S(x) \leq M$  for all  $x \in \mathbb{R}$ . To see this, observe that

$$(p_T * p_S)(x) = \int_{\mathbb{R}} p_T(y) p_S(x - y) dy \leq \int_{\mathbb{R}} p_T(y) M dy = M.$$

*Proof of (55):* From (18) in Lemma 1, we obtain that there exists  $M_1 \in \mathbb{N}$  such that for all  $m > M_1$  and  $n \in \mathbb{N}$  we have

$$\mathbb{E} \left[ \left| \tilde{A}_m \right|^\alpha \right] \leq 2\mathbb{E} [|A|^\alpha] + 1, \quad (58)$$

$$\mathbb{E} \left[ \left| \tilde{A}_{nm} \right|^\alpha \right] \leq 2\mathbb{E} [|A_n|^\alpha] + 1. \quad (59)$$

From (58) we obtain that

$$\mathbb{E} \left[ \left| \tilde{A}_m \right|^\alpha \right] \leq 2V + 1 < 2V + 3. \quad (60)$$

From (16) in Theorem 6 we obtain that there exists  $N_1 \in \mathbb{N}$  such that for all  $n > N_1$  we have

$$\mathbb{E} \left[ \left| \tilde{A}_n \right|^\alpha \right] \leq \mathbb{E} [|A|^\alpha] + 1. \quad (61)$$

Therefore, due to (59) and (61), we obtain that

$$m > M_1, n > N_1 \implies \mathbb{E} \left[ \left| \tilde{A}_{nm} \right|^\alpha \right] \leq 2V + 3,$$

Thus, (55) is proved.  $\square$

### C. Proof of Theorem 5

*Proof of (9):* Since  $\{(m_k, n_k)\}_{k=1}^\infty \in \text{FSP}_{\text{Stable}}^{\bar{\alpha}}$ , we conclude that  $\{(m_k, n_k)\}_{k=1}^\infty \in \text{FSP}_{\text{Stable}}^{\alpha_1} \cap \text{FSP}_{\text{Stable}}^{\alpha_2}$ . Hence, we can write

$$\begin{aligned} \mathcal{H}_{m_k, n_k}(X_1) - \mathcal{H}_{m_k, n_k}(X_2) &= n_k \left[ \frac{\mathcal{H}_{m_k, n_k}(X_1) - n_k \log \frac{m_k}{\alpha_1 \sqrt[n_k]{n_k}}}{n_k} \right. \\ &\quad \left. - \frac{\mathcal{H}_{m_k, n_k}(X_2) - n_k \log \frac{m_k}{\alpha_2 \sqrt[n_k]{n_k}}}{n_k} + \log \frac{\alpha_2 \sqrt[n_k]{n_k}}{\alpha_1 \sqrt[n_k]{n_k}} \right]. \end{aligned}$$

From Theorem 3 we obtain that there exists  $K$  such that for  $k > K$  we have

$$\frac{\mathcal{H}_{m_k, n_k}(X_1) - n_k \log \frac{m_k}{\alpha_1 \sqrt[n_k]{n_k}}}{n_k} - \frac{\mathcal{H}_{m_k, n_k}(X_2) - n_k \log \frac{m_k}{\alpha_2 \sqrt[n_k]{n_k}}}{n_k} \geq C,$$

where  $C := h(X_0^{(\alpha_1)}) - h(X_0^{(\alpha_2)}) - 1$ . Therefore, for  $k > K$ , we have that

$$\mathcal{H}_{m_k, n_k}(X_1) - \mathcal{H}_{m_k, n_k}(X_2) \geq n_k \left( C + \log \frac{\alpha_2 \sqrt[n_k]{n_k}}{\alpha_1 \sqrt[n_k]{n_k}} \right).$$

Hence,  $\mathcal{H}_{m_k, n_k}(X_1) - \mathcal{H}_{m_k, n_k}(X_2) \rightarrow +\infty$  as  $k \rightarrow \infty$ .

*Proof of (10):* We can write

$$\begin{aligned} \mathcal{H}_{m_k, n_k}(X_1) - \mathcal{H}_{m_k, n_k}(X_2) &= [\mathcal{H}_{m_k, n_k}(X_1) - \lambda_1 \log(m_k n_k)] - [\mathcal{H}_{m_k, n_k}(X_2) - \lambda_2 \log(m_k n_k)] \\ &\quad + (\lambda_1 - \lambda_2) \log(m_k n_k). \end{aligned}$$

From Theorem 4, we obtain that  $\mathcal{H}_{m_k, n_k}(X_1) - \mathcal{H}_{m_k, n_k}(X_2) \rightarrow +\infty$  as  $k \rightarrow \infty$ .

*Proof of (11):* We can write

$$\frac{\mathcal{H}_{m_k, n_k}(Y)}{\mathcal{H}_{m_k, n_k}(X)} = \frac{\log m_k n_k}{n_k \log \frac{m_k}{\sqrt[n_k]{n_k}}} \frac{\frac{\mathcal{H}_{m_k, n_k}(Y)}{\log m_k n_k}}{\frac{\mathcal{H}_{m_k, n_k}(X)}{n_k \log \frac{m_k}{\sqrt[n_k]{n_k}}}}$$

From Theorem 3 we obtain that

$$\lim_{k \rightarrow \infty} \frac{\mathcal{H}_{m_k, n_k}(X)}{n_k \log \frac{m_k}{\sqrt[n_k]{n_k}}} = 1,$$

since  $n_k, m_k / \sqrt[n_k]{n_k} \rightarrow \infty$  as  $k \rightarrow \infty$ . Furthermore, from Theorem 4 we obtain that

$$\lim_{k \rightarrow \infty} \frac{\mathcal{H}_{m_k, n_k}(Y)}{\log m_k n_k} = \lambda.$$

Therefore,

$$\lim_{k \rightarrow \infty} \frac{\mathcal{H}_{m_k, n_k}(Y)}{\mathcal{H}_{m_k, n_k}(X)} = \lambda \lim_{k \rightarrow \infty} \frac{\log m_k n_k}{n_k \log \frac{m_k}{\sqrt[n_k]{n_k}}} = \lambda \lim_{k \rightarrow \infty} \left[ \frac{\log m_k}{n_k} \frac{1}{\log \frac{m_k}{\sqrt[n_k]{n_k}}} + \frac{\log n_k}{n_k} \frac{1}{\log \frac{m_k}{\sqrt[n_k]{n_k}}} \right] = 0,$$

where the last equality is true since (i)  $(\log m_k)/n_k \rightarrow 0$  due to  $\{(m_k, n_k)\}_{k=1}^\infty \in \text{FSP}_{\text{Poisson}}$ , (ii)  $m_k / \sqrt[n_k]{n_k} \rightarrow \infty$  since  $\{(m_k, n_k)\}_{k=1}^\infty \in \text{FSP}_{\text{Stable}}^\alpha$ , and (iii) the fact that  $(\log n_k)/n_k \rightarrow 0$ .

Therefore, the theorem is proved.  $\square$

#### D. Proof of Theorem 6

*Proof of (13):* First, we find the characteristic function of  $Y_n$  in terms of the characteristic function of  $A$ . From the definition of  $Y_n$  in (12), we have that

$$\widehat{p}_{Y_n}(\omega) = \mathbb{E} [e^{j\omega Y_n}] = \mathbb{E} [e^{j\omega \langle X_n, \phi_{1,n} \rangle}],$$

where  $\phi_{1,n}$  defined in Definition 7. Due to the definition of white noise in Definition 10, we can write

$$\widehat{p}_{Y_n}(\omega) = \mathbb{E} [e^{j\langle X_n, \omega \phi_{1,n} \rangle}] = e^{\int_0^{\frac{1}{n}} f(\omega) dt} = e^{\frac{1}{n} f(\omega)}.$$

Because of the definition of Poisson process in Definition 13, we have that

$$f(\omega) = \lambda \int_{\mathbb{R}} (e^{j\omega x} - 1) p_A(x) dx = \lambda \widehat{p}_A(\omega) - \lambda,$$

where  $\widehat{p}_A(\omega)$  is the characteristic function of  $A$ . Hence, we conclude that

$$\begin{aligned}\widehat{p}_{Y_n}(\omega) &= e^{-\frac{\lambda}{n}} e^{\frac{\lambda}{n} \widehat{p}_A(\omega)} = e^{-\frac{\lambda}{n}} \left( 1 + \sum_{k=1}^{\infty} \frac{\left( \frac{\lambda}{n} \widehat{p}_A(\omega) \right)^k}{k!} \right) \\ &= e^{-\frac{\lambda}{n}} + \left( 1 - e^{-\frac{\lambda}{n}} \right) \widehat{p}_{A_n}(\omega),\end{aligned}\tag{62}$$

where

$$\widehat{p}_{A_n}(\omega) := \frac{1}{e^{\frac{\lambda}{n}} - 1} \sum_{k=1}^{\infty} \frac{\left( \frac{\lambda}{n} \widehat{p}_A(\omega) \right)^k}{k!}.\tag{63}$$

Therefore, by taking the inverse Fourier transform of (62), (13) is proved.

*Proof of (14):* It is achieved by taking the inverse Fourier transform from (63).

*Proof of (15):* We can write

$$\begin{aligned}p_{A_n}(x) &= \frac{1}{e^{\frac{\lambda}{n}} - 1} \sum_{k=1}^{\infty} \frac{\left( \frac{\lambda}{n} \right)^k}{k!} \left( \overbrace{p_A * \cdots * p_A}^k \right)(x) \\ &= \frac{\frac{\lambda}{n}}{e^{\frac{\lambda}{n}} - 1} p_A(x) + \frac{1}{e^{\frac{\lambda}{n}} - 1} \sum_{k=2}^{\infty} \frac{\left( \frac{\lambda}{n} \right)^k}{k!} \left( \overbrace{p_A * \cdots * p_A}^k \right)(x).\end{aligned}$$

Hence, we can write:

$$|p_{A_n}(x) - p_A(x)| \leq \left| \frac{\frac{\lambda}{n}}{e^{\frac{\lambda}{n}} - 1} - 1 \right| p_A(x) + \frac{1}{e^{\frac{\lambda}{n}} - 1} \sum_{k=2}^{\infty} \frac{\left( \frac{\lambda}{n} \right)^k}{k!} \left( \overbrace{p_A * \cdots * p_A}^k \right)(x).$$

Since,  $p_A$  is a probability density,  $\overbrace{p_A * \cdots * p_A}^k$  is also a probability density; hence,

$$\int_{\mathbb{R}} \left( \overbrace{p_A * \cdots * p_A}^k \right)(x) dx = 1.$$

Therefore,

$$\int_{\mathbb{R}} |p_{A_n}(x) - p_A(x)| dx \leq \left| \frac{\frac{\lambda}{n}}{e^{\frac{\lambda}{n}} - 1} - 1 \right| + \frac{1}{e^{\frac{\lambda}{n}} - 1} \sum_{k=2}^{\infty} \frac{\left( \frac{\lambda}{n} \right)^k}{k!} = 2 \frac{e^{\frac{\lambda}{n}} - \frac{\lambda}{n} - 1}{e^{\frac{\lambda}{n}} - 1},$$

which vanishes as  $n$  tends to  $\infty$ .

*Proof of (16):* Let the sequence of independent and identically distributed random variables  $\{A^{(i)}\}_{i=1}^{\infty}$  have the distribution  $p_A$ , and let the random variable  $B$  be independent to  $\{A^{(i)}\}_{i=1}^{\infty}$  and have the following distribution:

$$P_B[k] := \Pr \{B = k\} = \frac{1}{e^{\frac{\lambda}{n}} - 1} \frac{\left( \frac{\lambda}{n} \right)^k}{k!}, \quad \forall k \in \mathbb{N}.$$

From (14), we obtain that

$$A_n = \sum_{i=1}^B A^{(i)}, \quad (\text{in distribution}),$$

Now, to find  $\mathbb{E}[|A_n|^\alpha]$ , we can write

$$\mathbb{E}[|A_n|^\alpha] = \sum_{k=1}^{\infty} P_B[k] \mathbb{E} \left[ \left| \sum_{i=1}^k A^{(i)} \right|^\alpha \right] = \frac{\frac{\lambda}{n}}{e^{\frac{\lambda}{n}} - 1} \mathbb{E}[|A|^\alpha] + \frac{1}{e^{\frac{\lambda}{n}} - 1} \sum_{k=2}^{\infty} \frac{(\frac{\lambda}{n})^k}{k!} \mathbb{E} \left[ \left| \sum_{i=1}^k A^{(i)} \right|^\alpha \right].$$

From

$$\lim_{n \rightarrow \infty} \frac{\frac{\lambda}{n}}{e^{\frac{\lambda}{n}} - 1} = 1,$$

we have to prove that

$$\lim_{n \rightarrow \infty} \frac{1}{e^{\frac{\lambda}{n}} - 1} \sum_{k=2}^{\infty} \frac{(\frac{\lambda}{n})^k}{k!} \mathbb{E} \left[ \left| \sum_{i=1}^k A^{(i)} \right|^\alpha \right] = 0. \quad (64)$$

In order to prove (64), it suffices to show that

$$\mathbb{E} \left[ \left| \sum_{i=1}^k A^{(i)} \right|^\alpha \right] \leq c^k, \quad (65)$$

where  $c = 2^\alpha(1 + \mathbb{E}[|A|^\alpha])$  is a constant. This is because

$$\frac{1}{e^{\frac{\lambda}{n}} - 1} \sum_{k=2}^{\infty} \frac{(\frac{\lambda}{n})^k}{k!} \mathbb{E} \left[ \left| \sum_{i=1}^k A^{(i)} \right|^\alpha \right] \leq \frac{1}{e^{\frac{\lambda}{n}} - 1} \sum_{k=2}^{\infty} \frac{(\frac{\lambda}{n})^k}{k!} c^k = \frac{e^{\frac{\lambda}{n}c} - \frac{\lambda}{n}c - 1}{e^{\frac{\lambda}{n}} - 1},$$

vanishes as  $n$  tends to infinity. Therefore, it only remains to prove (65). We can bound  $|A^{(1)} + \dots + A^{(k)}|$  as follows:

$$\left| \sum_{i=1}^k A^{(i)} \right| \leq \sum_{i=1}^k |A^{(i)}| \leq 1 + \sum_{i=1}^k |A^{(i)}| \leq \prod_{i=1}^k (1 + |A^{(i)}|) \leq \prod_{i=1}^k (2 \max\{1, |A^{(i)}|\}).$$

By finding the expected value of the  $\alpha$  power of both sides, we have

$$\mathbb{E} \left[ \left| \sum_{i=1}^k A^{(i)} \right|^\alpha \right] \leq 2^{\alpha k} \mathbb{E} \left[ \prod_{i=1}^k \max\{1, |A^{(i)}|^\alpha\} \right] \quad (66)$$

$$= 2^{\alpha k} \prod_{i=1}^k \mathbb{E} \left[ \max\{1, |A^{(i)}|^\alpha\} \right] \quad (67)$$

$$= (2^\alpha \mathbb{E}[\max\{1, |A|^\alpha\}])^k, \quad (68)$$

where (66) is true because  $\max\{1, |x|\}^\alpha = \max\{1, |x|^\alpha\}$  for all  $x \in \mathbb{R}$ , (67) is true because  $A^{(1)}, \dots, A^{(n)}$  are independent, and (68) is true because  $A^{(1)}, \dots, A^{(n)}$  are identically distributed.

In order to find an upper bound for (68), we can write

$$\mathbb{E}[\max\{1, |A|^\alpha\}] \leq \mathbb{E}[1 + |A|^\alpha] = 1 + \mathbb{E}[|A|^\alpha]$$

Hence, (65), and the theorem are proved.  $\square$



### E. Proof of Lemma 1

*Proof of (17):* From the definition of  $q_{X;m}$  and  $q_{Y;m}$ , defined in Definition 6, it follows that they are constant in intervals  $[(i - 1/2)/m, (i + 1/2)/m]$  for all  $i \in \mathbb{Z}$ . Therefore,

$$\begin{aligned} \int_{\frac{i-\frac{1}{2}}{m}}^{\frac{i+\frac{1}{2}}{m}} |q_{Y;m}(x) - q_{X;m}(x)| \, dx &= \frac{1}{m} \left| q_{Y;m}\left(\frac{i}{m}\right) - q_{X;m}\left(\frac{i}{m}\right) \right| \, dx \\ &= \left| \int_{\frac{i-\frac{1}{2}}{m}}^{\frac{i+\frac{1}{2}}{m}} p_Y(x) \, dx - \int_{\frac{i-\frac{1}{2}}{m}}^{\frac{i+\frac{1}{2}}{m}} p_X(x) \, dx \right| \\ &\leq \int_{\frac{i-\frac{1}{2}}{m}}^{\frac{i+\frac{1}{2}}{m}} |p_Y(x) - p_X(x)| \, dx, \end{aligned} \quad (69)$$

where (69) is from the definition of  $q$ . Hence by summation over  $i \in \mathbb{Z}$ , we have

$$\int_{\mathbb{R}} |q_{Y;m}(x) - q_{X;m}(x)| \, dx \leq \int_{\mathbb{R}} |p_Y(x) - p_X(x)| \, dx.$$

Thus, (17) is proved.

*Proof of (18) and (19):* We claim that, it suffices to show that

$$|\tilde{X}_m - X| \leq \frac{1}{m}. \quad (70)$$

Because if so, we can write

$$||X| - \frac{1}{m}|^\alpha \leq |\tilde{X}_m|^\alpha \leq (|X| + \frac{1}{m})^\alpha,$$

and as a result

$$\mathbb{E} [||X| - \frac{1}{m}|^\alpha] \leq \mathbb{E} [|\tilde{X}_m|^\alpha] \leq \mathbb{E} [(|X| + \frac{1}{m})^\alpha].$$

Therefore, it suffices to show that

$$\mathbb{E} [(|X| + \frac{1}{m})^\alpha] \leq \left(\frac{2}{\sqrt{m}}\right)^\alpha + e^{\frac{\alpha}{\sqrt{m}}} \mathbb{E} [|X|^\alpha], \quad (71)$$

$$\mathbb{E} [||X| - \frac{1}{m}|^\alpha] \geq \Pr \left\{ |X| > \frac{1}{\sqrt{m}} \right\} e^{\frac{-2\alpha}{\sqrt{m}}} \mathbb{E} [|X|^\alpha]. \quad (72)$$

*Proof of (71):* By conditioning whether  $|X| > 1/\sqrt{m}$  or not, we can write

$$\begin{aligned} \mathbb{E} [(|X| + \frac{1}{m})^\alpha] &= \Pr \left\{ |X| > \frac{1}{\sqrt{m}} \right\} \mathbb{E} \left[ (|X| + \frac{1}{m})^\alpha \middle| |X| > \frac{1}{\sqrt{m}} \right] \\ &\quad + \Pr \left\{ |X| \leq \frac{1}{\sqrt{m}} \right\} \mathbb{E} \left[ (|X| + \frac{1}{m})^\alpha \middle| |X| \leq \frac{1}{\sqrt{m}} \right] \\ &\leq \Pr \left\{ |X| > \frac{1}{\sqrt{m}} \right\} \mathbb{E} \left[ (|X| + \frac{1}{m})^\alpha \middle| |X| > \frac{1}{\sqrt{m}} \right] + \left(\frac{2}{\sqrt{m}}\right)^\alpha, \end{aligned} \quad (73)$$

where (73) is true since

$$\Pr \left\{ |X| \leq \frac{1}{\sqrt{m}} \right\} < 1, \quad \frac{1}{m} + \frac{1}{\sqrt{m}} \leq \frac{2}{\sqrt{m}}.$$

Note that if  $\Pr \left\{ |X| > \frac{1}{\sqrt{m}} \right\} = 0$ , the conditional expected value  $\mathbb{E}[(|X| + \frac{1}{m})^\alpha | |X| > \frac{1}{\sqrt{m}}]$  is not well-defined, but in this case we take the product  $\Pr \left\{ |X| > \frac{1}{\sqrt{m}} \right\} \mathbb{E}[(|X| + \frac{1}{m})^\alpha | |X| > \frac{1}{\sqrt{m}}]$  to be zero, without any need for specifying an exact value for  $\mathbb{E}[(|X| + \frac{1}{m})^\alpha | |X| > \frac{1}{\sqrt{m}}]$ .

From (73), it only remains to prove that

$$\Pr \left\{ |X| > \frac{1}{\sqrt{m}} \right\} \mathbb{E} \left[ \left( |X| + \frac{1}{m} \right)^\alpha \middle| |X| > \frac{1}{\sqrt{m}} \right] \leq e^{\frac{\alpha}{\sqrt{m}}} \mathbb{E} [|X|^\alpha].$$

In order to do that, we can write

$$\left( |X| + \frac{1}{m} \right)^\alpha = |X|^\alpha e^{\alpha \log(1 + \frac{1}{m|X|})}.$$

Since,  $|X| > 1/\sqrt{m}$ , we can write

$$|X|^\alpha e^{\alpha \log(1 + \frac{1}{m|X|})} \leq |X|^\alpha e^{\alpha \log(1 + \frac{1}{\sqrt{m}})} \leq |X|^\alpha e^{\alpha \frac{1}{\sqrt{m}}},$$

where the last inequality is true because  $\log(1 + x) \leq x$ . As a result

$$\begin{aligned} \Pr \left\{ |X| > \frac{1}{\sqrt{m}} \right\} \mathbb{E} \left[ \left( |X| + \frac{1}{m} \right)^\alpha \middle| |X| > \frac{1}{\sqrt{m}} \right] &\leq \Pr \left\{ |X| > \frac{1}{\sqrt{m}} \right\} \mathbb{E} \left[ |X|^\alpha e^{\alpha \frac{1}{\sqrt{m}}} \middle| |X| > \frac{1}{\sqrt{m}} \right] \\ &\leq e^{\frac{\alpha}{\sqrt{m}}} \mathbb{E} [|X|^\alpha]. \end{aligned}$$

*Proof of (72):* Similar to the previous proof, by conditioning whether  $|X| > 1/\sqrt{m}$  or not, we can write

$$\begin{aligned} \mathbb{E} [|X| - \frac{1}{m}]^\alpha &= \Pr \left\{ |X| > \frac{1}{\sqrt{m}} \right\} \mathbb{E} \left[ \left| |X| - \frac{1}{m} \right|^\alpha \middle| |X| > \frac{1}{\sqrt{m}} \right] \\ &\quad + \Pr \left\{ |X| \leq \frac{1}{\sqrt{m}} \right\} \mathbb{E} \left[ \left| |X| - \frac{1}{m} \right|^\alpha \middle| |X| \leq \frac{1}{\sqrt{m}} \right] \\ &\geq \Pr \left\{ |X| > \frac{1}{\sqrt{m}} \right\} \mathbb{E} \left[ \left| |X| - \frac{1}{m} \right|^\alpha \middle| |X| > \frac{1}{\sqrt{m}} \right], \end{aligned} \tag{74}$$

If  $\Pr \left\{ |X| > 1/\sqrt{m} \right\} = 0$ , (72) is clearly correct. Thus, assume that  $\Pr \left\{ |X| > 1/\sqrt{m} \right\} > 0$ , meaning that  $\mathbb{E} \left[ \left| |X| - \frac{1}{m} \right|^\alpha \middle| |X| > 1/\sqrt{m} \right]$  is well-defined. We need to show that

$$\mathbb{E} \left[ \left| |X| - \frac{1}{m} \right|^\alpha \middle| |X| > \frac{1}{\sqrt{m}} \right] \geq e^{-\frac{\alpha}{\sqrt{m}}} \mathbb{E} [|X|^\alpha].$$

Observe that when  $|x| > 1/\sqrt{m}$  and  $m \geq 4$ , we can write

$$\left| |x| - \frac{1}{m} \right|^\alpha = |x|^\alpha e^{\alpha \log|1 - \frac{1}{m|x}|} \geq |x|^\alpha e^{\alpha \log(1 - \frac{1}{\sqrt{m}})} \geq |x|^\alpha e^{-\alpha \frac{2}{\sqrt{m}}},$$

where the last inequality is true because  $\log(1-x) \geq -2x$  for  $0 \leq x \leq 1/2$ . As a result,

$$\mathbb{E} \left[ \left| |X| - \frac{1}{m} \right|^\alpha \middle| |X| > \frac{1}{\sqrt{m}} \right] \geq e^{-\frac{\alpha}{\sqrt{m}}} \mathbb{E} \left[ |X|^\alpha \middle| |X| > \frac{1}{\sqrt{m}} \right].$$

Now, we need to show that

$$\mathbb{E} \left[ |X|^\alpha \middle| |X| > \frac{1}{\sqrt{m}} \right] \geq \mathbb{E} [|X|^\alpha]. \quad (75)$$

Without loss of generality, we may assume that  $\Pr \{|X| > 1/\sqrt{m}\} < 1$ . Now, we have

$$\begin{aligned} \mathbb{E} [|X|^\alpha] &= \Pr \left\{ |X| > \frac{1}{\sqrt{m}} \right\} \mathbb{E} \left[ |X|^\alpha \middle| |X| > \frac{1}{\sqrt{m}} \right] + \Pr \left\{ |X| \leq \frac{1}{\sqrt{m}} \right\} \mathbb{E} \left[ |X|^\alpha \middle| |X| \leq \frac{1}{\sqrt{m}} \right] \\ &\leq \max \left\{ \mathbb{E} \left[ |X|^\alpha \middle| |X| > \frac{1}{\sqrt{m}} \right], \mathbb{E} \left[ |X|^\alpha \middle| |X| \leq \frac{1}{\sqrt{m}} \right] \right\}. \end{aligned} \quad (76)$$

Since,

$$\mathbb{E} \left[ |X|^\alpha \middle| |X| \leq \frac{1}{\sqrt{m}} \right] \leq \left( \frac{1}{\sqrt{m}} \right)^\alpha < \mathbb{E} \left[ |X|^\alpha \middle| |X| > \frac{1}{\sqrt{m}} \right],$$

we obtain that

$$\max \left\{ \mathbb{E} \left[ |X|^\alpha \middle| |X| > \frac{1}{\sqrt{m}} \right], \mathbb{E} \left[ |X|^\alpha \middle| |X| \leq \frac{1}{\sqrt{m}} \right] \right\} = \mathbb{E} \left[ |X|^\alpha \middle| |X| > \frac{1}{\sqrt{m}} \right].$$

Therefore, (75) is obtained from (76). Hence, (72) is proved.

*Proof of (70):* Note that random variable  $\tilde{X}_m$  has the same distribution as the following random variable:

$$\tilde{X}_m = [X]_m + U_m,$$

where  $U_m$  and  $X$  are independent, and

$$p_{U_m}(x) = \begin{cases} m & |x| \leq \frac{1}{2m} \\ 0 & |x| > \frac{1}{2m} \end{cases}.$$

Thus,

$$|\tilde{X}_m - X| \leq |\tilde{X}_m - [X]_m| + |[X]_m - X| \leq |U_m| + \frac{1}{2m} \leq \frac{1}{m},$$

where the last inequality is true because  $|[X]_m - X|$  and  $|U_m|$  are always less than  $1/2m$ .

Therefore, the lemma is proved.  $\square$

### F. Proof of Lemma 2

From the definition of entropy, we know that the entropy of a random variable does not depend on the value of the random variable, rather it only depends on the distribution of the random variable. Therefore, if one finds a correspondence between the values of  $[aX]_{m/a}$  and the values of  $[X]_m$ , while they have the same probability, then the entropy of them will be the same. Thus, we define the following correspondence between the values of  $[aX]_{m/a}$ , which are from the set  $\{k \frac{a}{m} | k \in \mathbb{Z}\}$ , and the values of  $[X]_m$ , which are from the set  $\{k \frac{1}{m} | k \in \mathbb{Z}\}$ .

$$k \frac{a}{m} \in \{k \frac{a}{m} | k \in \mathbb{Z}\} \leftrightarrow k \frac{1}{m} \in \{k \frac{1}{m} | k \in \mathbb{Z}\}$$

Now, we show that the correspondent values have the same probability.

$$\begin{aligned} \Pr \left\{ [aX]_{\frac{m}{a}} = k \frac{a}{m} \right\} &= \Pr \left\{ aX \in \left[ \left(k - \frac{1}{2}\right) \frac{a}{m}, \left(k + \frac{1}{2}\right) \frac{a}{m} \right) \right\} \\ &= \Pr \left\{ X \in \left[ \left(k - \frac{1}{2}\right) \frac{1}{m}, \left(k + \frac{1}{2}\right) \frac{1}{m} \right) \right\} = \Pr \left\{ [X]_m = k \frac{1}{m} \right\}. \end{aligned}$$

Therefore, the lemma is proved.  $\square$

### G. Proof of Lemma 3

First, we show that it is only needed to prove that

$$\lim_{m \rightarrow \infty} H([X + d_m]_m) - \log m = h(X), \quad (77)$$

where  $\{d_m\}_{i=1}^{\infty}$  is an arbitrary sequence such that

$$d_m \in \left[0, \frac{1}{m}\right). \quad (78)$$

To show this, we know that for any given  $m, c_m \in \mathbb{R}$ , there exist unique  $k_m \in \mathbb{Z}$ , and  $d_m \in \mathbb{R}$  such that

$$c_m = \frac{k_m}{m} + d_m, \quad d_m \in \left[0, \frac{1}{m}\right).$$

Therefore, because of the definition of amplitude quantization in Definition 6, we can write

$$[X + c_m]_m = [X + d_m]_m + \frac{k_m}{m}.$$

Since entropy is invariant with respect to constant shift, we have

$$H([X + c_m]_m) = H([X + d_m]_m).$$

Thus, we proved the sufficiency of showing (77).

Theorem 2 (which is proved in [16]) is a special case of (77) when  $d_m = 0$  for all  $m$ . However, its proof can be adapted to establish (77). Take the probability distribution  $q_m(x)$  as follows:

$$q_m(x) = m \Pr \left\{ X + d_m \in \left[ \frac{i-\frac{1}{2}}{m}, \frac{i+\frac{1}{2}}{m} \right) \right\} = m \int_{\frac{i-\frac{1}{2}}{m}-d_m}^{\frac{i+\frac{1}{2}}{m}-d_m} p(x) dx, \quad (79)$$

when  $x \in [(i - \frac{1}{2})/m, (i + \frac{1}{2})/m)$ . Observe that

$$\begin{aligned} \int_{\frac{i-\frac{1}{2}}{m}}^{\frac{i+\frac{1}{2}}{m}} q_m(x) \log \frac{1}{q_m(x)} dx &= \frac{1}{m} q_m\left(\frac{i}{m}\right) \log \frac{1}{q_m\left(\frac{i}{m}\right)} \\ &= \frac{1}{m} q_m\left(\frac{i}{m}\right) \log \frac{1}{\frac{1}{m} q_m\left(\frac{i}{m}\right)} - \frac{1}{m} q_m\left(\frac{i}{m}\right) \log m. \end{aligned}$$

Therefore, by summing over  $i \in \mathbb{Z}$ , we obtain that

$$\int_{\mathbb{R}} q_m(x) \log \frac{1}{q_m(x)} dx = \sum_{i \in \mathbb{Z}} \frac{q_m(i/m)}{m} \log \frac{1}{\frac{1}{m} q_m(i/m)} - \log m = H([X + d_m]_m) - \log m,$$

where the last equation is true because of (79). So, if we take  $\tilde{X}_m \in \mathcal{AC}$  with distribution  $q_m$ , we only need to prove

$$\lim_{m \rightarrow \infty} h(\tilde{A}_m) = h(X).$$

In order to do this, we write  $|h(\tilde{A}_m) - h(X)|$  as follows

$$\begin{aligned} |h(\tilde{A}_m) - h(X)| &\leq \left| \int_{-l}^l q_m(x) \log \frac{1}{q_m(x)} - p(x) \log \frac{1}{p(x)} dx \right| + \left| \int_{|x|>l} q_m(x) \log \frac{1}{q_m(x)} dx \right| \\ &\quad + \left| \int_{|x|>l} p(x) \log \frac{1}{p(x)} dx \right|, \end{aligned}$$

where  $l > 0$  is arbitrary. Thus, it suffices to prove that

$$\lim_{m \rightarrow \infty} \int_{-l}^l q_m(x) \log \frac{1}{q_m(x)} dx = \int_{-l}^l p(x) \log \frac{1}{p(x)} dx, \quad (80)$$

$$\lim_{l \rightarrow \infty} \left| \int_{|x|>l} p(x) \log \frac{1}{p(x)} dx \right| = 0, \quad (81)$$

$$\lim_{l \rightarrow \infty} \left| \int_{|x|>l} q_m(x) \log \frac{1}{q_m(x)} dx \right| = 0, \quad (82)$$

hold for all  $l > 0$  and uniformly on  $m$ .

*Proof of (80):* From (79), we obtain that since  $p(x)$  is a piecewise continuous, the mean value theorem yields that there exists  $x_m^* \in [(i - \frac{1}{2})/m - d_m, (i + \frac{1}{2})/m - d_m)$ , such that  $q_m(x) = p(x_m^*)$ . Since  $p(x)$  is piecewise continuous, and  $[-l, l]$  is a compact set,  $p(x)$  is uniformly

continuous over  $[-l, l]$ . Therefore, for any  $\epsilon'$ , there exists  $M \in \mathbb{R}$  such that for all  $x \in [-l, l]$ , we have that

$$m > M \Rightarrow |q_m(x) - p(x)| < \epsilon'.$$

Furthermore, since the function  $x \mapsto x \log x$  is continuous and  $p(x)$  is uniformly continuous, we have that for all  $x \in [-l, l]$

$$\lim_{m \rightarrow \infty} q_m(x) \log \frac{1}{q_m(x)} = p(x) \log \frac{1}{p(x)},$$

uniformly on  $x$ . Thus (80) is proved.

*Proof of (81):* We can write

$$\left| \int_{|x|>l} p(x) \log \frac{1}{p(x)} dx \right| \leq \int_{|x|>l} p(x) \left| \log \frac{1}{p(x)} \right| dx. \quad (83)$$

By assuming the lemma, we know

$$\int_{\mathbb{R}} p(x) \left| \log \frac{1}{p(x)} \right| dx < \infty \Rightarrow \lim_{l \rightarrow \infty} \int_{|x|>l} p(x) \left| \log \frac{1}{p(x)} \right| dx = 0. \quad (84)$$

Hence, (83) implies (81).

*Proof of (82):* It suffices to show that

$$\int_{|x+d_m|>l} p(x) \log \frac{1}{p(x)} dx \leq \int_{|x|>l} q_m(x) \log \frac{1}{q_m(x)} dx, \quad (85)$$

$$\int_{|x|>l} q_m(x) \log \frac{1}{q_m(x)} dx \leq \sum_{|i|>l-1} P[i] \log \frac{1}{P[i]} + \Pr\{|X| > l-1\} \log 2 + o_l(1), \quad (86)$$

where  $o_l(1)$  means that  $\lim_{l \rightarrow \infty} o_l(1) = 0$ , and

$$P[i] = \int_{i-\frac{1}{2}}^{i+\frac{1}{2}} p(x) dx.$$

By changing the variable  $y = Lx$ , we can write

$$\begin{aligned} \int_{|x+d_m|>l} p(x) \log \frac{1}{p(x)} dx &= \int_{|x+d_m|>l} p(x) \log \frac{1}{\frac{1}{L}p(x)} dx - \Pr\{|X+d_m| > l\} \log L, \\ &= \int_{|y+Ld_m|>lL} \frac{1}{L}p\left(\frac{y}{L}\right) \log \frac{1}{\frac{1}{L}p\left(\frac{y}{L}\right)} dy - \Pr\{|X+d_m| > l\} \log L, \end{aligned}$$

Because  $p(x) < L$  almost everywhere, we have that:

$$\frac{1}{L}p\left(\frac{y}{L}\right) < 1 \Rightarrow \frac{1}{L}p\left(\frac{y}{L}\right) \log \frac{1}{\frac{1}{L}p\left(\frac{y}{L}\right)} > 0.$$

So we obtain that

$$\begin{aligned} \int_{|y+Ld_m|>lL} \frac{1}{L} p\left(\frac{y}{L}\right) \log \frac{1}{\frac{1}{L} p\left(\frac{y}{L}\right)} dy &\geq \int_{|y|>lL+L} \frac{1}{L} p\left(\frac{y}{L}\right) \log \frac{1}{\frac{1}{L} p\left(\frac{y}{L}\right)} dy \\ &= \int_{|x|>l+1} p(x) \log \frac{1}{p(x)} dx + \Pr\{|X| > l+1\} \log L, \end{aligned}$$

where  $x = y/L$ . Therefore, we can write that

$$\int_{|x+d_m|>l} p(x) \log \frac{1}{p(x)} dx \geq \int_{|x|>l+1} p(x) \log \frac{1}{p(x)} dx - \Pr\{|X| \in [l, l+1]\} \log L.$$

Thus, from (84) we conclude that the lower bound vanishes as  $l$  tends to  $\infty$  uniformly on  $m$ . Now, we are going to show that (86) leads that the upper bound vanishes as  $l$  tends to  $\infty$  uniformly on  $m$ . In order to prove this, note that  $\Pr\{|X| > l-1\}$  vanishes as  $l$  tends to infinity uniformly on  $m$ . Furthermore,

$$H([X]_1) < \infty \Rightarrow \sum_{i \in \mathbb{Z}} P[i] \log \frac{1}{P[i]} < \infty \Rightarrow \lim_{l \rightarrow \infty} \sum_{|i|>l-1} P[i] \log \frac{1}{P[i]} = 0.$$

Thus, in order to prove (82), it only remains to prove (85) and (86). The proof of (85) exists in [16] in the proof of Theorem 1. Thus, we only need to prove (86). Similar to the proof of Theorem 1 in [16], it can be shown that

$$\int_{|x|>l} q_m(x) \log \frac{1}{q_m(x)} dx \leq \sum_{|i|>l} P_m[i] \log \frac{1}{P_m[i]},$$

where  $P_m[i] := \Pr\{[X + d_m]_1 = i\}$ . This implies

$$H([X + d_m]_1 | |[X + d_m]_1| > l) = \sum_{|i|>l} \frac{P_m[i]}{\Pr\{|[X + d_m]_1| > l\}} \log \frac{\Pr\{|[X + d_m]_1| > l\}}{P_m[i]},$$

where  $H(X|Y = y)$  is defined in [22, p. 29]. Hence, we can write

$$\begin{aligned} \sum_{|i|>l} P_m[i] \log \frac{1}{P_m[i]} &= \Pr\{|[X + d_m]_1| > l\} H([X + d_m]_1 | |[X + d_m]_1| > l) \\ &\quad + \Pr\{|[X + d_m]_1| > l\} \log \frac{1}{\Pr\{|[X + d_m]_1| > l\}}. \end{aligned} \tag{87}$$

Because of the definition of the quantization, we can write

$$[X + d_m]_1 = [X]_1 + E_m,$$

where  $E_m$  is a random variable taking values from  $\{0, 1\}$ . Thus,  $[X + d_m]_1$  is a function of  $[X]_1$  and  $E_m$ ; as a result

$$\begin{aligned} H([X + d_m]_1 | [X + d_m]_1 > l) &\leq H([X]_1, E_m | [X + d_m]_1 > l) \\ &\leq H([X]_1 | [X + d_m]_1 > l) + H(E_m) \\ &\leq H([X]_1 | [X + d_m]_1 > l) + \log 2. \end{aligned}$$

From the definition of  $H([X]_1 | [X + d_m]_1 > l)$ , we can write that

$$\begin{aligned} &\Pr\{|[X + d_m]_1| > l\} H([X]_1 | [X + d_m]_1 > l) \\ &= \sum_{|i| > l} P[i] \log \frac{1}{P[i]} - r_m \log r_m - s_m \log s_m + t_m \log t_m \\ &\quad - \Pr\{|[X + d_m]_1| > l\} \log \frac{1}{\Pr\{|[X + d_m]_1| > l\}}, \end{aligned}$$

where

$$\begin{aligned} r_m &= \Pr\left\{X \in \left(l - \frac{1}{2} - d_m, l - \frac{1}{2}\right)\right\}, \quad s_m = \Pr\left\{X \in \left[-l - \frac{1}{2}, -l + \frac{1}{2} - d_m\right)\right\}, \\ t_m &= \Pr\left\{X \in \left[-l - \frac{1}{2}, -l + \frac{1}{2}\right)\right\}. \end{aligned}$$

Therefore, (87) can be simplified as follows:

$$\sum_{|i| > l} P_m[i] \log \frac{1}{P_m[i]} \leq \sum_{|i| > l} P[i] \log \frac{1}{P[i]} \quad (88)$$

$$- r_m \log r_m - s_m \log s_m + t_m \log t_m \quad (89)$$

$$+ \Pr\{|[X + d_m]_1| > l\} \log 2. \quad (90)$$

The first term, (88), vanishes as  $l$  tends to  $\infty$  because

$$H([X]_1) = \sum_{i \in \mathbb{Z}} P[i] \log \frac{1}{P[i]} < \infty \Rightarrow \lim_{l \rightarrow \infty} \sum_{|i| > l} P[i] \log \frac{1}{P[i]} = 0.$$

It can be achieved that (89) and (90) vanish as  $l$  tends to  $\infty$ , uniformly on  $m$ , but we do not write the details here. Thus, (86) is proved and the proof of lemma is complete.  $\square$

## VI. CONCLUSION

In this paper, a definition of entropy for random processes based on quantization in the time and amplitude domains was given. The criterion was applied to  $\alpha$ -stable and impulsive Poisson innovation processes, and it was shown that the  $\alpha$ -stable has a higher growth rate of entropy compared to the Poisson process. Characterization of the entropy for other Lévy processes is left as a future work.



## REFERENCES

- [1] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] E. Candès, “Compressive sampling,” in *Proc. International Congress of Mathematicians*, vol. 3, 2006, pp. 1433–1452.
- [3] R. Baraniuk, V. Cevher, and M. Wakin, “Low-dimensional models for dimensionality reduction and signal recovery: A geometric perspective,” *Proc. of the IEEE*, vol. 98, no. 6, pp. 959–971, 2010.
- [4] R. A. DeVore, “Nonlinear approximation,” *Acta numerica*, vol. 7, pp. 51–150, 1998.
- [5] A. Cohen, W. Dahmen, and R. DeVore, “Compressed sensing and best  $k$ -term approximation,” *J. Amer. Math. Soc.*, vol. 22, no. 1, pp. 211–231, 2009.
- [6] V. Cevher, “Learning with compressible priors,” in *Proc. Adv. in Neural Inform. Process. Sys. (NIPS)*, 2009, pp. 261–269.
- [7] A. Amini, M. Unser, and F. Marvasti, “Compressibility of deterministic and random infinite sequences,” *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5193–5201, Nov. 2011.
- [8] J. F. Silva and E. Pavez, “Compressibility of infinite sequences and its interplay with compressed sensing recovery,” in *Proc. Sig. & Inform. Proc. Assoc. Annual Summit and Conf. (APSIPA ASC), 2012 Asia-Pacific*, Dec. 2012.
- [9] R. Gribonval, V. Cevher, and M. E. Davies, “Compressible distributions for high-dimensional statistics,” *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5016–5034, 2012.
- [10] J. F. Silva and M. S. Derpich, “On the characterization of  $\ell_p$ -compressible ergodic sequences,” *IEEE Trans. Signal Process.*, vol. 63, no. 11, pp. 2915–2928, Jun. 2015.
- [11] M. Unser, P. Tafti, and Q. Sun, “A unified formulation of Gaussian versus sparse stochastic processes—Part I: Continuous-domain theory,” *IEEE Trans. Inf. Theory*, vol. 60, no. 3, pp. 1945–1962, 2014.
- [12] M. Unser, P. Tafti, A. Amini, and H. Kirshner, “A unified formulation of Gaussian versus sparse stochastic processes—Part II: Discrete-domain theory,” *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 3036–3051, 2014.
- [13] M. Unser and P. D. Tafti, *An introduction to sparse stochastic processes*. Cambridge University Press, 2014.
- [14] A. Amini, U. S. Kamilov, and M. Unser, “The analog formulation of sparsity implies infinite divisibility and rules out Bernoulli-Gaussian priors,” in *Inform. Theo. Workshop (ITW), 2012 IEEE*. IEEE, 2012, pp. 682–686.
- [15] A. Amini and M. Unser, “Sparsity and infinite divisibility,” *IEEE Trans. Inf. Theory*, vol. 60, no. 4, pp. 2346–2358, Apr. 2014.
- [16] A. Rényi, “On the dimension and entropy of probability distributions,” *Acta Mathematica Academiae Scientiarum Hungarica*, vol. 10, no. 1-2, pp. 193–215, Mar. 1959.
- [17] Y. Wu and S. Verdú, “Rényi information dimension: Fundamental limits of almost lossless analog compression,” *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3721–3748, Aug. 2010.
- [18] G. Lorentz, “Metric entropy and approximation,” *Bulletin of the American Mathematical Society*, vol. 72, no. 6, pp. 903–937, 1966.
- [19] E. C. Posner and E. R. Rodemich, “Epsilon entropy of stochastic processes with continuous paths,” *The Annals of Probability*, pp. 674–689, Aug. 1973.
- [20] S. Ihara, *Information theory for continuous systems*. Singapore: World Scientific, 1993.
- [21] C. Nair, B. Prabhakar, and D. Shah, “On entropy for mixtures of discrete and continuous variables,” *arXiv preprint cs/0607075*, 2006.
- [22] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. New York: John Wiley & Sons, 2006.
- [23] H. Ghourchian, A. Gohari, and A. Amini, “Existence and continuity of differential entropy for a class of distributions,” *accepted for publication in IEEE Communications Letters*, 2017.

- [24] G. Samoradnitsky and M. S. Taqqu, *Stable non-Gaussian random processes: stochastic models with infinite variance*. New York: Chapman & Hall, 1994.

## APPENDIX A

### WHITE LÉVY NOISES

To introduce the notion of white Lévy noises, we first define the concept of a *functional* that generalizes ordinary functions to include distributions such as Dirac's delta function and its derivatives [13].

**Definition 8** (Schwartz Space). *[13, p. 30] The Schwartz space, denoted as  $\mathcal{S}(\mathbb{R})$ , consists of infinitely differentiable functions  $\phi : \mathbb{R} \mapsto \mathbb{R}$ , for which*

$$\sup_{t \in \mathbb{R}} \left| t^m \frac{d^n}{dt^n} \phi(t) \right| < \infty, \quad \forall m, n \in \mathbb{N}.$$

*In other words,  $\mathcal{S}(\mathbb{R})$  is the class of smooth functions that, together with all of their derivatives, decay faster than the inverse of any polynomial at infinity.*

The space of *tempered distributions* or alternatively, the continuous dual of the Schwartz space denoted by  $\mathcal{S}'(\mathbb{R})$ , is the set of all continuous linear mappings from  $\mathcal{S}(\mathbb{R})$  into  $\mathbb{R}$  (also known as functionals). In other words, for all  $x \in \mathcal{S}'(\mathbb{R})$  and  $\varphi \in \mathcal{S}(\mathbb{R})$ ,  $x(\varphi)$  is a well-defined real number. Due to the linearity of the mapping with respect to  $\varphi$ , the following notations are interchangeably used:

$$x(\varphi) = \langle x, \varphi \rangle = \int x(t) \varphi(t) dt,$$

where  $\langle x, \varphi \rangle$ ,  $x(t)$  and the integral on the right-hand side are merely notations. This formalism is useful because it allows for a precise mathematical definition of functionals, such as impulse function that are common in engineering textbooks.

Just as functionals extend ordinary functions, generalized stochastic processes extend ordinary stochastic processes. In particular, a generalized stochastic process is a probability measure on  $\mathcal{S}'(\mathbb{R})$ . Further, observing a generalized stochastic process  $X(\cdot)$  is done by applying its realizations to Schwartz functions; *i.e.*, for a given  $\varphi \in \mathcal{S}(\mathbb{R})$ ,  $X_\varphi = \langle X, \varphi \rangle = \int X(t) \varphi(t) dt$  represents a real-valued random variable with

$$\Pr\{X_\varphi \in \mathcal{I}\} = \mu(\{x \in \mathcal{S}'(\mathbb{R}) \mid \langle x, \varphi \rangle \in \mathcal{I}\}),$$

$\mathcal{I} \in \mathcal{B}(\mathbb{R})$

where  $\mu(\cdot)$  stands for the probability measure on  $\mathcal{S}'(\mathbb{R})$  that defines the generalized stochastic process.

The white Lévy noises are a subclass of generalized stochastic processes with certain properties. Before we introduce them, we define Lévy exponents:

**Definition 9** (Lévy Exponent). [13, p. 59] A function  $f : \mathbb{R} \rightarrow \mathbb{C}$  is called a Lévy exponent if

- 1)  $f(0) = 0$ ,
- 2)  $f$  is continuous at 0,
- 3)  $\forall n \in \mathbb{N}, \forall \omega \in \mathbb{R}^n$ , and  $\forall \mathbf{a} \in \mathbb{C}^n$  satisfying  $\sum_{i=1}^n a_i = 0$ , we have that

$$\sum_{i,j=1}^n a_i a_j^* f(\omega_i - \omega_j) \geq 0.$$

The following theorem provides the algebraic characterization of Lévy exponents.

**Theorem 7** (Lévy-Khintchin). [13, p. 61] A function  $f(\omega)$  is a Lévy exponent if and only if it can be written as

$$-\frac{\sigma^2}{2}\omega^2 + j\mu\omega + \int_{\mathbb{R} \setminus \{0\}} (e^{j\omega a} - 1 - j\omega a \mathbb{1}_{|a|<1}(a)) v(a) da,$$

where  $\sigma, \mu \in \mathbb{R}$  are arbitrary constants. The function  $\mathbb{1}_{|a|<1}(a)$  is an indicator function which is 1 when  $|a| < 1$  and is 0 otherwise. The function  $v : \mathbb{R} \rightarrow [0, \infty)$  is a measure density function such that

$$\int_{\mathbb{R}} \min\{1, a^2\} v(a) da < \infty.$$

**Definition 10** (White Lévy Noises). [13, p. 73] A generalized stochastic process  $X$  is called a white Lévy noise, if

$$\mathbb{E} [e^{j\langle X, \varphi \rangle}] = \exp \left( \int_{\mathbb{R}} f(\varphi(t)) dt \right), \quad \forall \varphi \in \mathcal{S}(\mathbb{R}),$$

where  $\langle X, \varphi \rangle$  is the output of the linear operator  $\varphi$  under  $X$ , and  $f$  is a valid Lévy exponent.

The desired properties of a white noise could be inferred from Definition 10:

**Lemma 5.** [15] A white Lévy noise  $X$  is a stationary process in the sense that  $\langle X, \varphi_1 \rangle$  and  $\langle X, \varphi_2 \rangle$  have the same probability law when  $\varphi_2(t) = \varphi_1(t - t_0)$ . In addition, the independent atom property of white noise could be expressed as the statistical independence of  $\langle X, \varphi_1 \rangle$  and  $\langle X, \varphi_2 \rangle$  when  $\varphi_1$  and  $\varphi_2$  have disjoint supports ( $\varphi_1(t)\varphi_2(t) \equiv 0$ ).

Next, we explain two important types of white Lévy noises, namely stable and impulsive Poisson, that are studied in this paper.

**Definition 11** (Stable random variables). [24, p. 5] A random variable  $X$  is stable with parameters  $(\alpha, \beta, \sigma, \mu)$  if and only if its characteristic function  $\widehat{p}(\omega)$  is given by

$$\widehat{p}(\omega) := \mathbb{E} [e^{j\omega X}] = e^{f(\omega)},$$

where  $f : \mathbb{R} \mapsto \mathbb{C}$  is

$$f(\omega) = j\omega\mu - \sigma^\alpha |\omega|^\alpha (1 - j\beta \operatorname{sgn}(\omega) \Phi), \quad (91)$$

$$\Phi = \begin{cases} \tan \frac{\pi\alpha}{2}, & \alpha \neq 1, \\ -\frac{2}{\pi} \log |\omega|, & \alpha = 1, \end{cases}$$

$\alpha \in (0, 2]$  is the stability coefficient,  $\beta \in [-1, 1]$  is the skewness coefficient,  $\sigma \in (0, \infty)$  is the scale coefficient, and  $\mu \in \mathbb{R}$  is the shift coefficient. In addition, function  $\operatorname{sgn}(x) : \mathbb{R} \mapsto \{-1, 0, 1\}$  is the sign function defined as

$$\operatorname{sgn}(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}.$$

**Definition 12** (Stable white noise). [13, p. 64] The random process  $X$  is a stable innovation process with parameters  $(\alpha, \beta, \sigma, \mu)$  if  $X$  is a white Lévy noise with the Lévy exponent  $f(\omega)$  defined in (91).

**Definition 13** (Impulsive Poisson white noise). [13, p. 64] When the Lévy exponent  $f(\omega)$  of a white Lévy noise satisfies

$$f(\omega) = \lambda \int_{\mathbb{R} \setminus 0} (e^{ja\omega} - 1) p_A(a) da$$

for some scalar  $\lambda > 0$  (known as the rate of impulses) and probability density function  $p_A$  over  $\mathbb{R}$  (called the amplitude density), then, it is called an impulsive Poisson white noise.

## APPENDIX B

### PROOF OF LEMMA 4

*Proof of (20), (21), and (22):* If we prove that  $X_0$  is an stable random variable, then (20)-(22) are proved. [24] In order to do so, we find the characteristic function of  $X_0$ :

$$\widehat{p}_{X_0}(\omega) := \mathbb{E} [e^{j\omega \langle X, \phi \rangle}] = e^{\int_0^1 f(\omega) dt} = e^{f(\omega)}, \quad (92)$$

where,  $f(\omega)$  is a valid Lévy exponent, defined in (91), and (92) is true because of the definition of stable white noise in Definition 12. Hence, from Definition 11, we obtain that  $X_0$  is a stable random variable.

*Proof of (23):* Since, the functions  $x \mapsto p_{X_0}(x)$  and  $p \mapsto p|\log(1/p)|$  are continuous, the function  $x \mapsto p_{X_0}(x)|\log(1/p_{X_0}(x))|$  is also continuous. Thus, for any arbitrary  $x_0 > 0$ , we have

$$\int_{-x_0}^{+x_0} p_{X_0}(x) \left| \log \frac{1}{p_{X_0}(x)} \right| dx < \infty.$$

Therefore, in order to prove (23), we need to prove the boundedness of the tail of the integral. For sufficiently large  $x_0$ , we have

$$p_{X_0}(x) \leq \frac{c}{|x|^{\alpha+1}} \quad \forall x > x_0, \quad (93)$$

for some positive constant  $c$  depending on parameters  $(\alpha, \beta, \sigma, \mu)$  [24]. Since  $p \mapsto p|\log(1/p)|$  is increasing for  $p \in [0, 1/e]$ , it suffices to show that for sufficiently large  $x_0$ , the following integral is bounded:

$$\int_{|x|>x_0} \frac{c}{|x|^{\alpha+1}} \log \frac{|x|^{\alpha+1}}{c} dx < \infty. \quad (94)$$

By changing variable  $y = x/c^{1/(\alpha+1)}$ , it suffices to show that

$$\int_{|y|>y_0} \frac{1}{|y|^{\alpha+1}} \log |y|^{\alpha+1} dy < \infty, \quad (95)$$

where  $y_0 = x_0/c^{1/(\alpha+1)}$ , which holds.

*Proof of (24):* To show that  $H([X_0]_1) < \infty$ , it suffices to look at the tail of the probability sequence of  $[X_0]_1$ . From (93), for sufficiently large  $x_0$ , we have that for all  $|m| > x_0$

$$\int_{m-\frac{1}{2}}^{m+\frac{1}{2}} p_{X_0}(x) dx \leq \int_{m-\frac{1}{2}}^{m+\frac{1}{2}} \frac{a}{|x|^{\alpha+1}} dx < \frac{a}{(|m|-\frac{1}{2})^{\alpha+1}}. \quad (96)$$

Hence, it suffices to show that

$$\sum_{|m|-\frac{1}{2}>x_0} \frac{a}{(|m|-\frac{1}{2})^{\alpha+1}} \log \frac{(|m|-\frac{1}{2})^{\alpha+1}}{a} = 2 \sum_{m>x_0+\frac{1}{2}} \frac{a}{(m-\frac{1}{2})^{\alpha+1}} \log \frac{(m-\frac{1}{2})^{\alpha+1}}{a} < \infty.$$

due that  $a/x^{\alpha+1} \log(x^{\alpha+1}/a)$  is a decreasing function for sufficiently large  $x$ , we obtain that

$$\sum_{m \geq x_0+1} \frac{a}{(m-\frac{1}{2})^{\alpha+1}} \log \frac{(m-\frac{1}{2})^{\alpha+1}}{a} < \int_{x>x_0} \frac{a}{x^{\alpha+1}} \log \frac{x^{\alpha+1}}{a} dx < \infty,$$

where the last inequality is true because of (94).  $\square$